

Chatbots in the English Classroom: An Empirical Study of EFL Learner Interaction With a Written Dialog System

Dissertation

zur Erlangung der Würde einer Doktorin der Philosophie

vorgelegt dem

Institut für Bildungswissenschaften der Universität Basel

von

Johanna Oeschger

aus

Mettauertal

Biel / Bienne, 2024

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel

edoc.unibas.ch



Dieses Werk ist lizenziert unter einer

[Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).

Genehmigt vom Institut für Bildungswissenschaften der Universität Basel,
auf Antrag von Prof. Dr. Stefan Keller und Prof. Dr. Miriam Locher.

Basel, 18. Dezember 2023

Die Direktorin Prof. Dr. Elena Makarova

Abstract

Dialog systems are a promising technology to maximize opportunities for learner interaction in foreign language (FL) classrooms. Research has provided initial evidence for the effectiveness of learner-system interaction to promote FL development and for the occurrence of learning-facilitating interactional features, but in-depth investigations into the specific nature of FL learner-system interaction, particularly task-based interaction, are still outstanding. The present study explores the potential of dialog systems for FL learning further by describing the nature of task-based interaction between English as a foreign language (EFL) learners and a written dialog system compared to written synchronized computer-mediated communication (SCMC) learner-peer interaction. The study was situated in the Swiss commercial vocational education and training (VET) context where dialog systems may offer a particular benefit due to limited classroom time and self-organized computer-assisted learning.

The sample was collected in three commercial VET EFL classes using a written dialog system developed for the study based on a curriculum task (system group) and a written chat application (SCMC group), yielding 18 learner-system interactions and 21 learner-peer written SCMC learner-peer interactions. The two sets of interaction scripts were coded for interactional features and analyzed for group differences (system, SCMC) with respect to the learners' task performance in terms of complexity, accuracy, and communicative adequacy.

The findings revealed the occurrence of negotiation of meaning, recasts, and lexical and structural alignment in learner-system interaction at rates that were generally higher than in learner-peer SCMC and marked by characteristics unique to the learner-system interaction type. With regard to learners' task performance, the system group compared to the SCMC group produced language that was significantly less complex in terms of Analysis of Speech unit (AS-unit) length and lexical diversity measured by Measure of Textual Lexical Diversity (MTLD) and showed a non-significantly higher rate of inappropriate requests but was lexico-grammatically and orthographically non-significantly more accurate. Learners from the system group also

had a significantly lower task completion rate, but their generic structure and style ratings did not differ from those of the SCMC group.

Overall, the study's findings strengthen the case for the potential of dialog systems for certain aspects of FL learning, while also pointing to some important implications and limitations: Learner-system interaction may engage learners in frequent interactional features whose unique characteristics may be theoretically linked to FL learning functions and benefits as well as disadvantages. Compared to learner-peer SCMC, learner-system interaction may encourage learners to produce and practice language with higher lexico-grammatical and orthographic accuracy, but it may offer reduced opportunities for syntactically and lexically complex as well as pragmatically fully appropriate language use. Finally, learners interacting with a dialog system may be able to engage with the sociolinguistic and pragmatic competences targeted by the task's genre requirements in comparable ways as those interacting with a peer through SCMC; however, their ability to engage with the complete set of targeted competences might be reduced as they were more often unable to complete the task.

Acknowledgments

This is a revised version of my dissertation which was defended and accepted at the Institute for Educational Research at the University of Basel in 2023. The revisions were made in response to the feedback from my supervisors and consisted of restructuring and shortening the length of individual chapters, adding details to the documentation of methods, as well as minor formal corrections. The overall content of the dissertation remained unchanged.

Launching into this project at a time when ‘chatbots’ for language learning were still a relatively uncharted territory and I was already well into my career and life was an adventure that I could not have conquered without the support of many people around me.

I thank my supervisor Prof. Stefan Keller for venturing into this project with me and for his guidance and encouraging support throughout. Our discussions and his input have greatly contributed to my research and learning. I would also like to thank my supervisor Prof. Miriam Locher for her invaluable perspective and insights at the key stages of the project.

During the project I have had continued support and encouragement from my teachers and fellow Ph.D. students at FHNW School of Education and the University of Basel. I would like to especially thank Emily Raubach for her patient and diligent work as a second coder and Maleika Krüger, Cristina Vögelin, Charlotte Schneider, Eliane Gilg, Kirsten Schweinberger, and Ruth Trüb for their helpful advice and feedback on issues ranging from learner questionnaires and performance assessment to coding software and thesis formatting.

The project could not have taken off without the contribution of Kelvin Louis and Nicola Cocquio who developed the dialog system used in the study as part of their bachelor’s project with a lot of pioneering spirit, dedication, and hard work.

I would also like to thank the schools, students, and teachers who volunteered to participate in the study.

I am grateful for the generous financial support by the Aebli-Näf Foundation and my employer LerNetz who also granted me a reduced workload at key moments to enable me to move the project forward efficiently.

My family has been an unfaltering support for me throughout the project and I could not have done it without their help. Thank you to my husband, Ernie Deane, my parents, Ruth and Peter Oeschger, my siblings and their partners and children, Franziska and Oleg with Julia and Sascha Lavrovsky-Oeschger and Raphael Oeschger and Leslie Hansen with Paula Hansen, for being there through the highs and lows and for helping me find the time to work on my thesis by dedicating much of their own time for caring for our son. Thank you to our son, Jonah Oeschger, for being patient, for brightening up my life, and for putting everything into perspective.

Table of contents

1 Introduction.....	1
1.1 Purpose and Significance of the Study	7
1.2 Research Questions	8
1.3 Overview of the Study Design	8
1.4 Organization of the Dissertation	9
2 Literature Review	11
2.1 Cognitive-Interactionist Approach	11
2.1.1 Historical Origins	12
2.1.2 Theoretical Concepts and Empirical Evidence	16
2.1.3 Summary and Rationale for Choice of Theoretical Framework	42
2.2 Task-based Interaction	46
2.2.1 The Role of Task-based Interaction in FL Classroom Learning and Teaching	47
2.2.2 Evaluating Task Efficacy	50
2.3 SCMC for Language Learning	55
2.3.1 Claimed Benefits	56
2.3.2 Empirical Evidence	58
2.4 Dialog Systems for Language Learning	78
2.4.1 Claimed Benefits	81
2.4.2 Empirical Evidence	82
2.5 Research Gaps and Derivation of Research Questions	97
3 Methods	102
3.1 Research Design	103
3.2 Description and Rationale for Use of Mixed Methods Approach	105
3.3 Pilot Study 1: Task Development	107
3.4 Dialog System	109
3.4.1 Overview	111
3.4.2 NLU	112
3.4.3 Dialog Manager	114
3.4.4 Training and Testing	117
3.4.5 Usability Tests	120
3.5 Pilot Study 2: Infrastructure, Questionnaires, and Interview	122
3.6 Pilot Study 3: Final Questionnaires and Tasks	128
3.7 Main Study	129
3.7.1 Participants	129
3.7.2 Materials	132
3.7.3 Infrastructure	133
3.7.4 Procedure	134
3.7.5 Data Processing and Analysis	136
4 Findings	180
4.1 General Overview of Interactions	180
4.2 Research Question 1: Interactional Features	182

4.2.1 Negotiation of Meaning	183
4.2.2 Recasts.....	236
4.2.3 Lexical Alignment	261
4.2.4 Structural Alignment.....	290
4.2.5 Summary of Findings on Research Question 1.....	296
4.3 Research Question 2: Learners' Task Performance.....	303
4.3.1 Complexity	304
4.3.2 Accuracy.....	306
4.3.3 Communicative Adequacy	312
4.3.4 Summary of Findings on Research Question 2.....	322
5 Discussion.....	326
5.1 General Overview of Interactions	326
5.2 Research Question 1	327
5.2.1 Negotiation of Meaning	327
5.2.2 Recasts	338
5.2.3 Lexical Alignment.....	343
5.2.4 Structural Alignment	353
5.3 Research Question 2.....	359
5.3.1 Complexity.....	359
5.3.2 Accuracy.....	363
5.3.3 Communicative Adequacy.....	370
6 Conclusion.....	378
6.1 Review of the Study Purpose.....	378
6.2 Key Findings.....	378
6.2.1 Interactional Features	378
6.2.2 Learners' Task Performance.....	381
6.3 Implications for Teaching.....	384
6.3.1 Benefits and Challenges of Learner-System Interactional Features	384
6.3.2 Benefits and Challenges for Task-based Interaction	387
6.4 Limitations of the Current Study	388
6.5 Contributions and Suggestions for Further Research	392
6.5.1 Contributions.....	393
6.5.2 Suggestions for Further Research.....	393
6.6 Future of Dialog Systems for Language Learning: Potential and Risks	395
References.....	402
Appendices.....	433
Appendix A Topics and objectives of participating schools' English curriculum.....	433
Appendix B Communicative functions in coursebook sample texts matching dialog system's actions	434
Appendix C Task handouts	435
System group	435
SCMC group (learner/client role)	436
SCMC group (interlocutor/hotel employee role)	437
Appendix D Learner questionnaires.....	439
Pre-questionnaire.....	439

Post-questionnaire.....	443
Appendix E Example interview guide	452
Appendix F Test administration scripts	458
Week 1.....	458
Week 2.....	459
Week 3.....	461
Appendix G Student consent form.....	464
Appendix H Codebooks derived from interaction scripts	465
Appendix I Complete intercoder agreement measures	466
Appendix J Grouped responses to questions in post-questionnaire	474
Appendix K Codes pertaining to negotiation of meaning derived from learner interviews	476
Appendix L Codes pertaining to complexity derived from learner interviews	480
Appendix M Codes pertaining to accuracy derived from learner interviews	485
Appendix N Codes pertaining to communicative adequacy derived from learner interviews	490

List of Figures

Figure 1 <i>Visualization of the Dialog System's NLU Pipeline</i>	112
Figure 2 <i>Structured Representation of User Message by Dialog System</i>	113
Figure 3 <i>Visualization of the Dialog System's Dialog Manager Pipeline</i>	114
Figure 4 <i>Excerpt from List of Intents Derived from the Hotel Task Tested in Pilot 1</i>	118
Figure 5 <i>Excerpt from List of Entities Derived from the Hotel Task Tested in Pilot 1</i>	118
Figure 6 <i>Excerpt from the DSL File Used to Generate Alternative Formulations of Intents</i>	119
Figure 7 <i>Example of a Sample Conversation Used for the Training of the Dialog Manager</i>	120
Figure 8 <i>Classification of Intents by Dialog System in Usability Tests 1 – 3</i>	121
Figure 9 <i>Distribution of Dialog System Utterances Classified by Correctness and Type of Fallback Message</i>	122
Figure 10 <i>Participants' Frequency of Writing and Chatting in English</i>	131
Figure 11 <i>Participants' Frequency of Communicating with Dialog Systems in English or Any Other Language</i>	132
Figure 12 <i>User Interface of the Dialog System Application</i>	133
Figure 13 <i>User Interface of the SCMC Application</i>	134
Figure 14 <i>Visualization of Possible Paths Following an Indicator of Non-understanding in Negotiation of Meaning in Learner-System Interaction</i>	199
Figure 15 <i>Visualization of Possible Paths Following the Response Component in Negotiation of Meaning in Learner-System Interaction</i>	216
Figure 16 <i>Visualization of Possible Paths Following the Response Component in Negotiation of Meaning in SCMC</i>	226
Figure 17 <i>Code Matrix of Subcodes for Emotional Reactions to Negotiation of Meaning per Learner in Learner-System Interaction</i>	232
Figure 18 <i>Code Matrix of Subcodes for Conceptions on Negotiation of Meaning per Learner in Learner-System Interaction</i>	234

Figure 19 <i>Code Matrix of Subcodes for Strategies and Conceptions on Negotiation of Meaning per Learner</i>	236
Figure 20 <i>Distribution of Questionnaire Responses to Questions on Aspects of Communicative Adequacy</i>	316
Figure 21 <i>Code Matrix of Subcodes for Task Completion per Learner in Learner-System Interaction</i>	320

List of Tables

Table 1 <i>Overview of Research Design for Main Study</i>	103
Table 2 <i>Participants' Language Levels</i>	130
Table 3 <i>Overview of Main and Subcodes for Negotiation of Meaning</i>	140
Table 4 <i>Overview of Main and Subcodes for Recasts</i>	142
Table 5 <i>Overview of Subcodes for Lexical Alignment</i>	146
Table 6 <i>List of Frequent Syntactical Structures for Coding of Structural Alignment</i>	148
Table 7 <i>Overview of Main and Subcodes for Requests</i>	161
Table 8 <i>Overview of Measures of Complexity and Accuracy</i>	168
Table 9 <i>Subcodes/Levels and Corresponding Scores for the Rating of Communicative Adequacy</i>	170
Table 10 <i>Intercoder Agreement</i>	175
Table 11 <i>Descriptive Statistics and Results of Independent Sample T-Test (Two- tailed) and Mann-Whitney U Test for Differences in General Measures Between Learner-System Interaction and SCMC</i>	182
Table 12 <i>Frequency of Indicators of Negotiation of Meaning in Learner-System Interaction</i>	184
Table 13 <i>Overview of Main and Subcodes for Negotiation of Meaning with Counts</i>	185
Table 14 <i>Frequency and Distribution of Subcodes for Initiator of Negotiation of Meaning in Learner-System Interaction</i>	186
Table 15 <i>Frequency and Distribution of Subcodes for Type of Indicator of Negotiation of Meaning in Learner-System Interaction</i>	187
Table 16 <i>Frequency and Distribution of Subcodes for Accuracy of Trigger of Negotiation of Meaning in Learner-System Interaction</i>	195
Table 17 <i>Frequency and Distribution of Subcodes for Type of Response of Negotiation of Meaning in Learner-System Interaction</i>	199
Table 18 <i>Frequency and Distribution of Subcodes for Accuracy of Response in Negotiation of Meaning in Learner-System Interaction</i>	210

Table 19	<i>Frequency and Distribution of Subcodes for Type of Response Leading to a Resolution vs. No Resolution of the Non-Understanding in Learner-System Interaction</i>	212
Table 20	<i>Frequency and Distribution of Subcodes for Type of Reaction to Response of Negotiation of Meaning in Learner-System Interaction</i>	213
Table 21	<i>Frequency and Distribution of Resolved and Abandoned Sequences by Number of Layers of Negotiation of Meaning in Learner-System Interaction</i>	221
Table 22	<i>Frequency of Negotiation of Meaning Indicators in Learner-System Interaction and SCMC</i>	222
Table 23	<i>Frequency and Distribution of Subcodes for Type of Indicator of Negotiation of Meaning in Learner-System Interaction and SCMC</i>	223
Table 24	<i>Frequency and Distribution of Subcodes for Type of Indicator of Negotiation of Meaning in Learner-System Interaction and SCMC</i>	223
Table 25	<i>Frequency and Distribution of Subcodes for Accuracy of Trigger of Negotiation of Meaning in Learner-System Interaction and SCMC</i>	225
Table 26	<i>Frequency of Recasts in Learner-System Interaction</i>	237
Table 27	<i>Overview of Main and Subcodes for Recasts With Counts</i>	237
Table 28	<i>Frequency and Distribution of Subcodes for Type of Linguistic Focus of Recasts in Learner-System Interaction</i>	239
Table 29	<i>Frequency and Distribution of Subcodes for Intensity of Recasts Within Interaction in Learner-System Interaction</i>	244
Table 30	<i>Frequency and Distribution of Subcodes for Incidence of Target-Like Form Within Recast in Learner-System Interaction</i>	245
Table 31	<i>Frequency and Distribution of Subcodes for Completeness of Recasts in Learner-System Interaction</i>	246
Table 32	<i>Frequency and Distribution of Subcodes for Length of Recasts in Learner-System Interaction</i>	247
Table 33	<i>Frequency and Distribution of Subcodes for Uptake of Recasts in Learner-System Interaction</i>	249
Table 34	<i>Frequency and Distribution of Subcodes for Re-Uptake of Recasts in Learner-System Interaction</i>	251

Table 35	<i>Frequency of Recasts in Learner-System Interaction and SCMC</i>	253
Table 36	<i>Frequency and Distribution of Subcodes for Linguistic Focus of Recasts in Learner-System Interaction and SCMC</i>	254
Table 37	<i>Frequency and Distribution of Subcodes for Intensity of Recast Within Interaction in Learner-System Interaction and SCMC</i>	258
Table 38	<i>Frequency and Distribution of Subcodes for Intensity of Recast Form Within Recast in Learner-System Interaction and SCMC</i>	259
Table 39	<i>Frequency and Distribution of Subcodes for Completeness of Recasts in Learner-System Interaction and SCMC</i>	259
Table 40	<i>Frequency and Distribution of Subcodes for Length of Recast in Learner-System Interaction and SCMC</i>	260
Table 41	<i>Frequency and Distribution of Subcodes for Uptake of Recasts in Learner-System Interaction and SCMC</i>	260
Table 42	<i>Frequency and Distribution of Subcodes for Re-Uptake of Recasts in Learner-System Interaction and SCMC</i>	261
Table 43	<i>Frequency of Subcodes for Lexical Alignment in Learner-System Interaction</i>	263
Table 44	<i>Frequency of Lexical Alignment in Learner-System Interaction and SCMC</i>	280
Table 45	<i>Frequencies of Syntactical Structures in Learner-System Interaction and SCMC</i>	291
Table 46	<i>Frequency of Subcodes for Structural Alignment Across all Structures in Learner-System Interaction</i>	292
Table 47	<i>Frequencies of Subcodes for Structural Alignment Across all Structures in Learner-System Interaction and SCMC</i>	295
Table 48	<i>Descriptive Statistics and Results of Independent Sample T-Test (Two-tailed) for Differences in Complexity Measures Between Learner-System Interaction and SCMC</i>	305
Table 49	<i>Descriptive Statistics and Results of Independent Sample T-Test (Two-tailed) and Mann-Whitney U Test for Differences in Accuracy Measures Between Learner-System Interaction and SCMC</i>	307

Table 50 <i>Frequency and Distribution of Subcodes for Types of Errors in Learner-System Interaction and SCMC</i>	308
Table 51 <i>Frequency and Distribution of Subcodes/Levels for Appropriateness Ratings of Requests in Learner-System Interaction and SCMC</i>	308
Table 52 <i>Frequency and Distribution of Subcodes for Type of Request Strategy in Learner-System Interaction and SCMC</i>	309
Table 53 <i>Frequency and Distribution of Subcodes for Modifiers of Requests in Learner-System Interaction and SCMC</i>	311
Table 54 <i>Descriptive Statistics and Results of Mann-Whitney U Test for Differences in Communicative Adequacy Measures Between Learner-System Interaction and SCMC</i>	313
Table 55 <i>Frequency and Distribution of Subcodes/Levels of Communicative Adequacy in Learner-System Interaction and SCMC</i>	314

1 Introduction

Anecdotally, foreign language (FL) learners often state that they ‘really’ learned a language when using it in conversation. So, to FL learners – and their teachers – an interlocutor who is available at any time, never grows tired of a conversation, patiently listens to the learner’s formulation attempts, and repeats themselves without complaints as many times as needed sounds like an ideal conversational partner.

It is no surprise, then, that FL teachers and researchers have taken a keen interest in dialog systems such as chatbots that simulate a human interlocutor with the added benefit of being always available, having infinite patience and offering a low-anxiety environment to language learners (Bibauw et al. 2022). The research interest in dialog systems for FL learning and teaching has intensified in recent years as related technologies such as machine learning, natural language processing, and text recognition have improved, leading to dialog systems such as chatbots or virtual assistants being widely used in areas ranging from healthcare through to home entertainment and education (Bibauw et al., 2022; Fryer et al., 2020).

In the Swiss educational context, the commercial vocational education and training (VET)¹ program provides a particularly fertile ground to explore the potential of dialog systems in FL classrooms: On the one hand, the commercial VET program has the highest number of learners among all Swiss VET programs (Staatssekretariat für Bildung, Forschung und Innovation [SBFI], 2022) and thus constitutes a highly relevant context in the Swiss educational system. On the other hand, the commercial VET curriculum specifically asks for efficient (language) learning that is increasingly self-directed by learners, including by using digital devices and resources, as will be argued in more detail in the following paragraphs.

Due to Switzerland’s many internationally operating companies, the commercial VET program attaches high importance to FL learning, as stated in the foreign language concept by the Swiss Conference of Commercial Training and

¹ VET is a post-compulsory educational offer at upper-secondary level chosen by around two thirds of young people in Switzerland (SBFI, 2022). VET programs combine an apprenticeship in a training company with classroom learning to develop the skills and competencies that the learners need in the chosen profession. Successful completion of a three- or four-year VET program leads to a Federal VET Diploma.

Examination Branches (SKKAB) (2021b), the supporting association responsible for the development of the commercial VET program. The Federal Ordinance (SBFI, 2012) that regulates the implementation of the commercial VET program by the individual Swiss cantons and VET commercial schools, has therefore established, in general², two FLs as compulsory school subjects for commercial apprentices. The revised federal ordinance that has been in place since school year 2023/2024 (SBFI, 2021) has established one FL as a compulsory subject for all learners and a second FL as an elective subject. While the federal ordinance leaves it up to the cantons which FLs are to be taught at their commercial schools, due to its role as lingua franca in the globalized business world, English is generally considered particularly significant for commercial apprentices. This is reflected in the fact that the majority of commercial VET schools teach English as the main FL besides the national languages French, German, and Italian (Vindayer & Hodel, 2007).

The FL curriculum for the commercial VET program requires learners to develop a wide range of profession-specific language competences that match the demands of the commercial work environment. The curriculum that was in place at the time of this study (SKKAB, 2020) defines service encounters as most frequent communicative situations where the learner as service-seeker demands services and goods and requests information on them or, in the role of service provider, gives information and assistance about their own company's services and products to customers. Other objectives involve receptive and productive skills in the context of, for example, internal meetings and trainings or job applications.

However, limited classroom time narrows the scope of FL classroom teaching and learning in the commercial VET program. The curriculum in place at the time of the present study dedicated 240³ of a total of 1800 school lessons per year to each foreign language taught, limiting classroom time to two lessons per week⁴ (SKKAB,

² At the time when the present study was conducted and documented from 2018 to 2023, the federal ordinance allowed for a «basic» (B) as well as an «extended» (E) profile (SBFI, 2012, p. 2). While the E profile required learners to be taught in two compulsory FLs, learners in the B profile were required to learn only one FL. The revised federal ordinance that has been in place since 2023, has removed the option of choosing between two profiles, establishing one FL as a compulsory subject for all learners and a second FL as an elective subject.

³ This number of lessons applies to the E profile which teaches two FLs. The B profile teaching only one FL assigns 320 lessons to FL teaching to compensate for the different number of total lessons.

⁴ Cantons may choose to conduct the foreign language final examinations at the end of the fourth

2020). Limited classroom time typically restricts opportunities for learner interaction and, in particular, learner-teacher interaction as it is a time-consuming activity. Dialog systems promise an efficient and easily organizable resource for these FL classrooms to facilitate interaction despite time constraints.

The emphasis that the commercial VET program puts on self-organized learning is another reason why dialog system-interaction may be a useful learning tool for learners, as it allows for conversation independent of other students' or the teacher's availabilities. The commercial VET curriculum highlights that commercial employees require a high degree of self-reliance among other personal competencies, such as creativity, willingness for lifelong learning, and responsibility for sustainable action due to the profession's constant evolution in response to developments in technology, operational procedures, and new ways of collaborating with clients and partners (SKKAB, 2020). Accordingly, the curriculum also aims at social and personal competencies such as commitment or communication skills besides professional competencies and method competencies which are all to be developed as part of the professional competencies and in the course of existing school subjects. The revised curriculum in effect from 2023 continues the integration of these interdisciplinary competencies (SKKAB, 2021a). Consequently, VET commercial schools strive to integrate learner-directed activities as part of the existing school subjects that specifically help them to develop autonomy and personal responsibility. For instance, the VET commercial school where this study was conducted runs two classes in which learners study autonomously in a self-organized learning environment (Wirtschafts- und Kaderschule KV Bern [WKS], n.d.). In self-organized learning, the computer and online resources become important learning tools to learners as the teachers' role shifts to that of a "coach" (WKS, n.d.).

From the point of view of FL learning and teaching methodology, maximizing opportunities for learners to interact, including through new technology, is essential as interaction is considered fundamental to FL learning. The fundamental role of interaction for language learning is theorized by the cognitive-interactionist approach

semester of the three-year program, thus completing the curriculum in two rather than three years while increasing the weekly lessons from two to three lessons.

which posits that through interaction, FL learners receive comprehensible input, are pushed to produce output, and engage in interactional work such as negotiation of meaning, feedback, or alignment which promote the development of linguistic means, i.e., vocabulary, grammar, pronunciation (and orthography in written interaction) (Schocker & Müller-Hartmann, 2016). Robust evidence from studies conducted over the last three decades (A. Mackey et al., 2011) support the claims by the cognitive-interactionist approach both in terms of the overall benefits of interaction on FL development (e.g., meta-analyses by Keck et al., 2006; Mackey & Goo, 2007) as well as the benefits of its individual core components comprehensible and interactionally modified input (Ellis et al., 1994; Gass & Varonis, 1994; Loschky, 1994; A. Mackey, 1999; Pica et al., 1987), output (DeKeyser, 1997; Egi, 2010; Izumi et al., 1999; McDonough, 2005; M. Sato & Lyster, 2012; R. Sato, 2012; R. Sato & Takatsuka, 2016; Shehadeh, 2001), and interactional features such as negotiation of meaning (Bitchener, 2004; De la Fuente, 2002; Ellis et al., 1994; Nobuyoshi & Ellis, 1993; Pica, 1996), feedback (e.g., meta-analyses by Li, 2010; Lyster & Saito, 2010), and lexical and structural alignment (Dao et al., 2018; McDonough et al., 2015; McDonough & Mackey, 2008; Trofimovich et al., 2014).

Since school-based FL learning including commercial VET FL learning is based on tasks, the potential of dialog systems for classroom interaction must be evaluated in the context of task-based interaction. In task-based interaction learners interact on a task, which in FL learning research is defined as a meaning-focused activity that requires learners to achieve a goal in a context that is relevant to them, enabling them to use language authentically. Moreover, tasks focus on learners as agents and allow them to develop their linguistic competencies by focusing on linguistic form (Thonhauser, 2016). These characteristics make tasks and task-based interaction suitable for VET language classrooms for two reasons: First, tasks engage learners in meaning-making, and thus promote the interactional processes such as, for example, comprehensible output and negotiation of meaning that are linked to the development of linguistic means, as posited by the cognitive-interactionist approach (Ellis, 2003). Second, by placing learners in authentic communicative contexts, tasks require learners to activate their sociolinguistic and pragmatic skills, besides their linguistic competencies. Tasks thus play a key role in the development of communicative

competence which combines linguistic, sociolinguistic, and pragmatic competence (Council of Europe, 2001a). Communicative competence, in turn, is the goal of the communicative and competency-based approach that builds the foundation of VET FL learning and teaching: Following the Common European Framework of Reference for Languages (CEFR), the commercial VET FL curriculum formulates four core competencies for listening and speaking, reading, writing, and language foundations, each of which combines the three aspects of communicative competence, i.e., linguistic, sociolinguistic, and pragmatic competencies. Task-based interaction thus enables learners to develop communicative competence as required by the commercial VET FL curriculum.

Task-based interaction requires a type of dialog system that places no constraints on form and allows the learner to freely produce any kind of language input while implicitly constraining meaning by restricting its responses to the thematic scope of the task. This type of dialog system is referred to in the literature as a goal-oriented system with contextualized dialogue (Bibauw et al., 2019). The existing research on task-based learner interaction with goal-oriented dialog systems is still very limited. There is, however, substantial evidence on the efficacy of the modally similar task-based interaction mediated by synchronized computer-mediated communication (SCMC), i.e., chat, which has led to the hypothesis that dialog systems may offer comparable opportunities for interaction and FL learning (Bibauw et al., 2022).

Various studies have investigated whether the claims and concepts of the cognitive-interactionist approach also apply to SCMC. Some studies found that interactional features also occur in SCMC, including feedback (S. J. Sauro, 2013; S. M. Sotillo, 2000), negotiation of meaning (R. Blake, 2000; L. Lee, 2001; Smith, 2003), and alignment (Michel & Cappellini, 2019; Smith & Michel, 2017; Uzum, 2010). Other studies have provided evidence for the efficacy of the occurring features with regard to L2 development (e.g., Kim et al., 2020 for alignment; Smith, 2004 for negotiation of meaning). Furthermore, it was found that in some cases, the features' characteristics differ from those in face-to-face (FTF) communication (e.g., Smith, 2003 on negotiation of meaning) and that some of these SCMC-specific characteristics might in fact be more advantageous (e.g., Yuksel & Inan, 2014 for the role of negotiation of

meaning in noticing). There is also direct evidence for the positive effect of SCMC on L2 development (see the meta-analyses by Lin et al., 2013; Ziegler, 2016).

In addition to these findings on the presence, nature, and efficacy of interactional features in SCMC, research has also produced evidence on the usefulness of SCMC for task-based interaction, specifically. The efficacy of tasks – computer-mediated or not – is commonly evaluated by measuring the learners' performance on the task in terms of linguistic complexity, accuracy, and fluency as well as communicative adequacy, i.e., the degree to which the learner succeeds in efficiently achieving the tasks' goals (Pallotti, 2009, p. 596) considering aspects such as pragmatic appropriateness and successful task completion. SCMC studies on task performance have found equally or more complex and accurate language in SCMC compared to FTF interaction (Fitze, 2006; S. Sauro, 2012). Other studies provided evidence for the beneficial role of SCMC in the development of pragmatically appropriate production of address forms (Belz & Kinginger, 2002; E. Y. A. Kim & Brown, 2014; Tudini, 2007), discourse functions (Chun, 1994; R. G. Kern, 1995), and speech acts (Cunningham, 2016a; Sykes, 2005). Overall, while SCMC has not been researched as thoroughly as FTF interaction, the findings indicate that learners receive similar opportunities for the production of interactional features and development of their FL/L2 in SCMC and FTF contexts.

When considering the potential of dialog systems, these findings from SCMC research give rise to the question whether the modally similar dialog system-interaction offers learners the same benefits in terms of the cognitive-interactionist and task-based approach. While research on interactional features in learner-system L2 interaction is still very limited, initial findings are promising. Two studies found negotiation of meaning in learner-system interaction (N.-Y. Kim, 2017b; Satar & Qinghua, 2020) while another study provided evidence for the occurrence of alignment in learner-system interaction (Sinclair et al., 2019). These findings demonstrate that learner-system interaction displays some of the features of human-human interaction associated with L2 learning. In addition, there were indications that the interactional features of learner-system interaction might differ from those found in human-human interaction in terms of frequency (N.-Y. Kim, 2017b) as well as characteristics (Sinclair et al., 2019). Furthermore, research to date suggests that the interactional features in

learner-system interaction impact the effects of interaction on L2 development in similar ways as FTF interaction (Satar & Qinghua, 2020; Wilske, 2015). In terms of the effect on learners' task performance, learner-system interaction has been found to lead to at least the same quantity of text as well as variety of different words as a non-interactive form of written production (Nagata et al., 2020). Learners interacting with dialog system also seem to use some of the same request strategies and modifications as previously observed in FTF interaction, although there was a tendency towards more routinized, direct, and explicit requests (Timpe-Laughlin & Dombi, 2020). There is additional direct evidence for the positive impact of learner-system interaction on FL development, for example, from the first meta-analysis on the topic by Bibauw et al. (2022) who found an overall medium effect ($d = 0.59$) for learner-system interaction on FL development.

1.1 Purpose and Significance of the Study

The purpose of this study is to describe the nature of task-based interaction between commercial EFL learners and a dialog system with the goal to further explore its potential for commercial EFL classroom learning and teaching, specifically in the context of the commercial VET program where maximizing opportunities for learner interaction, including with new technology, is central due to limited classroom time and a self-organized learning approach.

As becomes evident from the brief overview above, more research on task-based learner-system interaction is needed to evaluate its potential for EFL classroom learning. A first research gap concerns the investigation of interactional features. This is an important step for the evaluation of dialog systems as, on the theoretical basis of the cognitive-interactionist approach, the occurrence of interactional features in learner-system interaction could be considered indirect evidence for the efficacy of learner-system interaction and a possible explanation for its positive effects, while analyzing the features' specific characteristics could point to possible advantages and challenges compared to human-human interaction.

A second research gap is the missing assessment of learners' task performance, and specifically learners' sociolinguistic and pragmatic performance when interacting

with a dialog system. Assessing learners' task performance not only in terms of linguistic measures is crucial as learners' language may be accurate and complex but functionally inefficient (Pallotti, 2009). Evaluating sociolinguistic and pragmatic aspects of task performance as measured by the construct of *communicative adequacy* (Pallotti, 2009) may indicate to what extent learners activate their pragmatic and sociolinguistic competencies besides their linguistic competencies when trying to achieve the task's intended communicative goal.

This study aims to close these gaps of knowledge by describing interactional features and task performance in task-based interaction between learners and a dialog system. It is hoped that this exploratory work will be useful to both FL learning research as well as teaching practice. It is aimed to contribute to the understanding of learner-system interaction, providing a foundation for future research. At the same time, it aims to inform FL teachers on the potential benefits and challenges as well as practical implications of using dialog systems in their FL classrooms.

1.2 Research Questions

The study is guided by the following research questions:

1. *What interactional features occur in written task-based interaction between commercial VET EFL learners and a dialog system and what are their characteristics compared to those found in learner-peer written SCMC on the same task?*
2. *Is there a difference in the learners' task performance in terms of linguistic complexity and accuracy as well as communicative adequacy between learner-system interaction and learner-peer SCMC?*

1.3 Overview of the Study Design

To answer the research questions, a written goal-oriented dialog system was developed to enable learner-system interaction on a task based on the commercial VET EFL curriculum. Learner-peer SCMC on the same task was carried out using the chat application *Mattermost*. A written rather than spoken or multimodal (i.e., written,

spoken, as well as visual, gestural, and haptic) dialog system was chosen, as written chat constitutes an unintrusive technology and is a familiar mode of communication for young Swiss learners, which makes it well-suited for the VET classroom setting.

In the resulting scripts, interactional features were qualitatively and quantitatively analyzed, and the learners' task performance was assessed in terms of complexity and accuracy as well as communicative adequacy. Retrospective interviews and post-questionnaires served to complement the findings and interpretation of the main data by adding the learners' perspective.

1.4 Organization of the Dissertation

The thesis is organized as follows: Chapter 1 introduced the topic and presented the rationale for carrying out the study as well as its significance for the field of EFL teaching and learning. The chapter concludes with a statement of the research questions and a brief overview of the study design.

Chapter 2 presents a review of the literature in four sections. The first section introduces the cognitive-interactionist approach to FL learning, along with its core concepts and related empirical research supporting the approach. The section concludes with a rationale for choosing the cognitive-interactionist approach as theoretical framework for this study. The second section discusses the central role of task-based interaction in the educational context of this study, followed by a description of how task-based interaction is evaluated by means of the concepts *complexity*, *accuracy*, *fluency* and *communicative competence*. The third and fourth sections present the research from the cognitive-interactionist and task-based approach on written SCMC and learner-system interaction, respectively. The final section highlights the gaps in the current state of research and restates and describes in detail the derived research questions.

Chapter 3 describes the methodology that was applied in the study. It first presents an overview and rationale of the chosen research design. It then reports on pilot study 1 for the development of the task used for data collection. This follows a description of the dialog system that was developed for this study and a brief report of its development process. The following two sections document pilot studies 2 and 3

which informed the selection of infrastructure (e.g., chat application), development of materials (e.g., questionnaires, interview guides), and procedure of the main study. The final section reports on the main study: It first presents the participants, materials, and data collection procedure, followed by a description of the data processing and analysis of the scripts. Finally, the section describes the analysis of the supplementary data gained in learner questionnaires and interviews.

Chapter 4 reports the findings related to the two research questions. It first presents a comparative general overview of the two interaction types, i.e., learner-system interaction and learner-peer SCMC. Then, the findings on research question 1 focusing on the characteristics of interactional features of learner-system interaction are presented, giving a detailed account of each feature's characteristics, including a side-to-side comparison with the features found in learner-peer SCMC as well as the insights gained from the learner interviews and questionnaire. The section on research question 2 presents the results from the statistical analyses testing for differences in task performance between the two groups as well as the findings from the learner interviews and questionnaire.

Chapter 5 summarizes the major findings with regard to the two research questions and interprets them by integrating the findings from the different data sources (i.e., interaction scripts, questionnaires, interviews) and relating them back to the theory and previous research. The findings are also discussed in terms of their implications for FL learning research and teaching.

Chapter 6 reviews the key points of the study considering the study purpose and discusses the study's limitations as well as its implications for future research and teaching. The chapter, and thesis, concludes with a reflection on the potential and risks of dialog systems for FL learning in the present and future.

2 Literature Review

This chapter reviews the literature relevant to the topic of the present study, task-based learner-system interaction for FL learning and teaching. The purpose of the chapter is to present the theoretical and empirical work that grounded the study. More specifically, the chapter serves to introduce the theoretical concepts and methodologies that informed the design, methods, and analysis of the study and to establish the current state of research in the field of human-human interaction (FTF, SCMC) for FL learning, which provided a basis for the rationale of investigating the use of dialog systems for FL learning.

The first section of the chapter introduces the cognitive-interactionist approach as the theoretical framework that guided the study, including an overview of its core concepts and empirical evidence supporting its claims. This is followed by a section on the literature explaining the central role of task-based interaction in classroom FL learning and teaching and introducing the approaches to evaluate the effectiveness of task-based interaction. The third section presents the empirical research on written SCMC, which informed this study in two ways: First, it framed the analysis of the SCMC data which was collected in addition to learner-system interaction data as part of the study design, as well as the analysis of the modally similar written learner-system interaction. Second, the empirical evidence on the efficacy of written SCMC provided a rationale for investigating the efficacy of the modally similar dialog system-interaction. The fourth section of the chapter presents the empirical research on dialog system interaction for FL learning. This is followed by a final section that summarizes and highlights the gaps in the current state of research and concludes by deriving and restating the research questions.

2.1 Cognitive-Interactionist Approach

The cognitive-interactionist approach explains the process of FL and second language (L2) acquisition by claiming that there is a strong connection between interaction and language development. In FL and L2 research, interaction denotes the conversation in which FL and L2 learners engage (Y. Kim, 2017). In FL and L2

classrooms, interaction typically takes place between learners and teacher or as peer interaction among learners during pair or group work. FL interaction may be carried out in person, also referred to as FTF interaction, or it can be mediated by communication technology (Y. Kim, 2017). Technology-mediated interaction may involve, on the one hand, computer-mediated communication (CMC) enabling learners to interact with another language user such as a peer or a native speaker through a computer. CMC is divided into synchronous computer-mediated communication (SCMC) (e.g., chat) and asynchronous computer-mediated communication (ACMC) (e.g., email) (R. Blake, 2017). On the other hand, technology-mediated interaction includes dialog system-interaction where learners interact with the computer as a virtual interlocutor (Bibauw et al., 2019). Both CMC and dialog system-interaction may be conducted in oral (e.g., audio or video chat), written (e.g., text chat), or multimodal modes (e.g., video chat combined with text chat) (Ziegler & Phung, 2019). The present study investigates the interaction between EFL learners and a written dialog system, comparing it to written learner-peer SCMC.

The present section outlines in a brief historical overview how the cognitive-interactionist approach grew out of different hypotheses established since the 1980s, which theorized the relationship between interaction and FL acquisition. The presented hypotheses also provide the foundation for the understanding of the core concepts of the cognitive-interactionist approach which will be introduced in the following section along with the empirical evidence on their role for FL acquisition. Finally, the section presents a rationale for the choice of the cognitive-interactionist approach as theoretical framework in this study.

2.1.1 Historical Origins

It is commonly agreed that the cognitive-interactionist approach has its origins in Long's interaction hypothesis (Bowles & Adams, 2015; Y. Kim, 2017; Lambert et al., 2019; Loewen & Sato, 2018). In his initial version of the interaction hypothesis, Long (1981, 1983) posits that the conversational modifications which occur when native speakers and non-native speakers work together to resolve a communication problem

make language input more comprehensible to learners, thus facilitating language acquisition. In a series of studies on learner-native speaker interaction, Long had observed that the interlocutors used a number of conversational strategies and modifications to avoid or deal with communication problems (e.g., treating topics briefly, selecting salient topics, requesting clarification, and tolerating ambiguity). Long argued that these strategies and modifications made input comprehensible to learners by supplying them with semantic information on linguistic forms that they had trouble understanding.

Long's focus on the central role of strategies and modifications to facilitate comprehensible input was highly influenced by previous research on foreigner talk, i.e., the modifications made by native speakers when talking to non-native speakers to make their language more comprehensible, while the claim that comprehensible input was a requirement for FL and L2 acquisition was influenced by Krashen's input hypothesis, developed in the 1970s and 1980s (Y. Kim, 2017; A. Mackey et al., 2011). Krashen's input hypothesis posited that for L2 and FL acquisition to take place, learners need to be exposed to input whose meaning they understand but that contains structures that are "beyond" their "current level of competence" (Krashen, 1982, p. 21), referred to by Krashen as stage "i + 1". To enable comprehension of language that is beyond the learners' current level of competence, their language input should be simplified (i.e., by using a slower speaking rate and clearer articulation) and enriched with contextual and extralinguistic clues (e.g., providing pictures, using it in the context of familiar topics).

However, the claim of the initial version of the interaction hypothesis about the sufficient role of comprehensible input for second and foreign acquisition was subsequently contested by a number of researchers (Mackey et al., 2011). Swain, for instance, responded by proposing the comprehensible output hypothesis (1985, 1995, 2005), which assumed a much more central role to the learner output in the process of FL and L2 acquisition, while, previously, output had been seen mainly as a source of "(self)-input" for learners (Swain, 1995, p. 125). Swain's view was motivated by empirical studies showing that comprehensible input was not sufficient for the development of certain aspects of language competence (e.g., higher order grammatical structures, pragmatic competence). Swain suspected that limited opportunity for output, i.e., to

produce language, was the cause. She argued that learners “can fake it, so to speak in comprehension, but they cannot do so in the same way in production” (Swain, 1995, p. 127). In other words, while in comprehension learners can rely on semantic processing, avoiding the processing of syntax, production forces learners to process linguistic forms bottom-up when attempting to produce accurate and appropriate output. The concept that learners are pushed in their production in order to make themselves understood is referred to by Swain “pushed” output (Swain, 1985, p. 249).

Another theoretical claim that emerged in the context of interaction following Long’s first version of the interaction hypothesis was Schmidt’s noticing hypothesis. According to Schmidt (1990), learners must “notice”, i.e., become consciously aware, of a feature in the input in order for it to become internalized. When learners process input at the lower awareness level of “perception” learning will be unsuccessful. However, Schmidt (2001) later modified this position, proposing that while attended learning is superior, some learning is possible without it, albeit “not much” (p. 40).

The theoretical claims made by Swain’s output hypothesis and Schmidt’s noticing hypothesis were subsequently subsumed in Long’s (1996) revised version of the interaction hypothesis:

Negotiation for meaning, and especially negotiation work that triggers interactional adjustment by the NS or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways. (pp. 451–452)

As becomes evident from this quote, the updated interaction hypothesis places particular importance on negotiation of meaning as a type of interactional modification in which the interlocutors engage when communication problems arise. As Long (1996) specifies, negotiation of meaning refer to episodes in which the interlocutor signals or the speaker perceives a comprehension issue, and the speaker adjusts their speech at the linguistic, content, or discourse level or a combination of all three in order to overcome the communication problem. And it is this negotiation work, as the updated interaction hypothesis claims, that facilitates acquisition as it provides learners with opportunities to receive comprehensible input and produce modified output as the mental mechanisms involved in the cognitive acquisition process are activated (Lambert et al., 2019).

These claims made by the interaction hypothesis have since been investigated by a large number of studies. Early research of the 1980s and 1990s worked to provide evidence to the claims of Long's initial interaction hypothesis. Researchers conducted descriptive studies (e.g., Doughty & Pica, 1986; Pica et al., 1987, 1993; Varonis & Gass, 1985) on interactions in learner-native speaker or peer dyads showing the high use of interactional modifications and comparing the effects of premodified and interactionally modified input on comprehension (A. Mackey et al., 2011; Mayo & Soler, 2013).

Later research on the updated version of the interaction hypothesis shifted its focus from investigating interactional modifications and the relationship between interactional modifications and comprehensible input to the role of these components in facilitating acquisition (Y. Kim, 2017; Loewen & Sato, 2018). Researchers also began investigating and providing positive evidence for the beneficial effect of interaction as a whole on FL and L2 development, such as, for example, on lexical acquisition (Ellis et al., 1994), overall performance (Gass & Varonis, 1994), or question formation (A. Mackey, 1999). More recently, meta-analyses have provided further evidence for an overall medium-to-large effect of interaction on acquisition (Keck et al., 2006; A. G. Mackey & Goo, 2007).

Based on the clear empirical support for the benefits of interaction, Mackey, & Gass (2007) concluded that "it is now commonly accepted within the SLA literature that there is a robust connection between interaction and learning" (p. 176), pointing to the agreement in the field that the hypothesis has evolved into a theoretical approach to FL and L2 learning (Ellis et al., 1994).

In sum, rooted in the interaction hypothesis, the cognitive-interactionist approach views interaction as a beneficial environment for FL and L2 learning as it provides learners with opportunities to receive comprehensible input, produce output, and be exposed to interactional features such as negotiation of meaning or corrective feedback which are involved in the mental mechanisms promoting acquisition. The following subsections describe each of the components (i.e., input, output, interactional features) and their functioning in more detail. For each feature, the empirical research corroborating these hypothetical claims will also be discussed.

2.1.2 Theoretical Concepts and Empirical Evidence

This subsection presents the theoretical concepts of the cognitive-interactionist approach by introducing each component of interaction (i.e., input, output, interactional features) along with their hypothesized benefits and empirical research.

As a note, the reviewed research includes studies on FTF interaction only, while the research on CMC and dialog system-interaction will be discussed in separate sections of the literature review. This structure reflects the chronological and argumentative order of the cognitive-interactionist research, which initially explored the theoretical claims in the context of FTF interaction, where they had emerged, before investigating their application to CMC and, more recently, dialog system-interaction.

Input

Input refers to the written and oral language that learners are exposed to during interaction (Gass, 2015; Y. Kim, 2017; Mayo & Soler, 2013). From the cognitive-interactionist perspective, input is considered a key component of FL and L2 acquisition as it provides learners with the linguistic data for FL/L2 processing. More precisely, input serves as positive evidence for learners to use as the basis for forming hypotheses about the FL/L2 (Gass, 2015; Y. Kim, 2017; Ziegler, 2018).

As explained above (see section Historical Origins), the cognitive-interactionist perspective holds that input must be comprehensible to the learner in order to facilitate acquisition. Comprehensible input may be achieved through modification, which can be done interactionally (i.e., through negotiation of meaning when the learner signals a comprehension problem) or it can be pre-modified by simplifying or enhancing a written or oral text (Han et al., 2008; Lambert et al., 2019). While a simplified text uses shorter sentences or decreased lexical complexity, elaborated input enhances the language e.g., by clarifying its meaning or increasing the saliency of certain target structures without simplifying it (Mayo & Soler, 2013). Unmodified input is referred to as “naturalistic” (Loewen & Sato, 2018, p. 287) or “authentic” input (Mayo & Soler, 2013, p. 214).

The cognitive-interactionist approach sees interactionally modified input as essential – whereas the input hypothesis highlights the role of pre-modified input to generate comprehensible input – as it allows for complex input thanks to overall conversational adjustments while supporting comprehensibility (Long, 1996; Mayo & Soler, 2013).

The cognitive-interactionist approach further suggests that the function of input relies on the learners' noticing of linguistic features in the input, which is achieved through the learners' focal attention to the linguistic input with some level of awareness (Gass, 2015; Y. Kim, 2017; Loewen & Sato, 2018). It has been suggested that attention to linguistic forms in the input may be achieved through enhancing the saliency of target structures by increasing their frequency or by highlighting them, e.g., by using intonation and stress in oral texts or typographic enhancement in written texts (Lambert et al., 2019). However, as Gass (2015) points out, the speaker's intention may not match the learner's perception.

The role of input for FL and L2 learning has been the focus of early cognitive-interactionist research seeking to find evidence for the positive impact of, on the one hand, input modifications on the comprehensibility of the input and, on the other hand, of comprehensible input on acquisition. The body of research that investigated the interactionally modified input specifically in terms of negotiation of meaning, will be reported below in the subsection dedicated to negotiation of meaning (see section Negotiation of Meaning).

Studies focusing on general input modifications and comprehensibility found that interactionally modified input was more successful in promoting comprehension compared to unmodified input (Pica et al., 1987) and also advantageous over pre-modified input (Ellis et al., 1994; Gass & Varonis, 1994; Loschky, 1994).

Studies on the impact of interactionally modified input on acquisition provided evidence for its beneficial effect: For instance, Ellis et al. (1994) compared the effects of pre-modified and interactionally modified input on comprehension and the acquisition of vocabulary in two separate studies, reporting an advantage of interactionally modified input with respect to both comprehension and the acquisition of the target items, at least in the short term. Similarly, Mackey (1999) explored the

relationship of interactional modifications and grammar development operationalized as the acquisition of English as a second language (ESL) question formation, i.e., the production of syntactically more advanced question forms. The findings indicated that the participants who actively interacted with their interlocutors achieved the greatest gains in L2 development compared to those that merely observed the interaction or received pre-modified input.

Output

In the context of FL and L2 interaction, output refers to the written and spoken language that learners produce in conversation (Y. Kim, 2017; Loewen & Sato, 2018). While it may seem trivial to state that speaking and writing a language play a crucial role in learning that language, it needs to be understood that prior to the establishment of the output hypothesis (see section Historical Origins) output was seen not as a way of creating new knowledge but as a manifestation of what a learner already knows (Ellis, 2008). Largely based on Swain's output hypothesis, the cognitive-interactionist approach suggests several ways in which output facilitates language learning: First, output may enhance fluency by providing learners with opportunities to practice productive performance of the FL/L2 (Swain, 1995). Fluency relies on the automatization of language use, i.e., the process of development from conscious and controlled declarative knowledge to automatized knowledge that can be retrieved more efficiently and correctly (DeKeyser et al., 2013). This suggests that in order to attain automatic and fluent language use, learners should have numerous and repeated opportunities to practice it.

Second, when attempting to produce the language, learners will become aware of their linguistic shortcomings, either through internal feedback generated by monitoring their own production or through external feedback by the interlocutor. Thus, they may realize that their output is not target-like – referred to as *notice the gap* (R. Schmidt & Frota, 1986) – or that they fail to express what they intend to convey – i.e., *notice the hole* (Swain 1995). This realization may prompt learners to direct their attention to the relevant linguistic information in the input which in turn triggers cognitive processes involved in FL/L2 development (Izumi, 2003).

As an additional benefit, learners may produce modified output in response to negative feedback (external or through monitoring) where they modify their initial output to make it more target-like. Modified output is considered a type of uptake, which in turn denotes the learner's utterance following and containing a reaction to feedback (Egi, 2010). The modified output learners produce in response to external feedback by an interlocutor, e.g., in the course of negotiation of meaning or corrective feedback, is referred to as *other-initiated* modified output. Modified output resulting from internal feedback is *self-initiated*.

Modified output does not necessarily correct the original error: Uptake that is unsuccessful in correcting the error of the initial utterance is referred to as *no repair* or *needs repair* (Lyster & Ranta, 1997). Uptake that successfully corrects the error is referred to as *repair* and partially successful correction of the error as *partial repair*. The successful result of self-initiated modified output is referred to as *self-repair* (Ellis, 2008).

With respect to the benefits of modified output and repair, it is suggested that not only repair but also the process of modifying output itself, is beneficial to language development, regardless of whether learners produce the correct target forms or not (Ziegler & Bryfonski, 2018), as it can be seen as surface manifestation of the cognitive processes involved in FL/L2 acquisition as well as a mechanism for FL/L2 learning (Ellis, 2008).

Third, output has a hypothesis-testing function: When producing output, learners can judge the comprehensibility and linguistic adequacy of their production against their interlocutor's feedback (Swain, 1995).

Finally, the learners' output may serve a metalinguistic function by allowing learners to "reflect upon their own target language use" through language (Swain, 1995, pp. 126–127), thus deepening the learners' language awareness (Izumi, 2003).

The claims on these roles of output for FL and L2 acquisition have been empirically researched in a number of studies. With regard to its function in the development of automatization and fluency, an early longitudinal study by DeKeyser (1997), for instance, investigated the differential effect of comprehension and production practice on the automatization of L2 grammar items. The study found that

both practice types gradually led to increased automatization, measured by a drop in the number of errors and reaction times. Another finding was that the effect of practice was skill-specific, i.e., learners who only practiced comprehension improved their comprehension but far less their production performance, whereas learners from the production group improved their performance but less so their comprehension performance. These findings suggest that input alone is not sufficient for L2 acquisition of both comprehension and production competence but that learners need to be provided opportunities to produce output as well in order to improve both skills.

Sato and Lyster (2012) provided further evidence for the role of output, specifically modified output as a reaction to feedback, for L2 development in terms of fluency and accuracy. The study explored the differential effects of peer interaction and two corrective feedback types (prompts and recasts) embedded in peer interaction on the English development of Japanese learners. Language gains in the three experimental groups were assessed against a control group who had no extra interaction opportunities outside the regular curriculum activities after one semester of intervention in terms of accuracy and oral fluency. The authors found that all three interaction types led to improved fluency, which they ascribe to the experimental groups' repeated opportunities for output production. Additionally, interaction enhanced with corrective feedback had a positive impact on accuracy development. According to the authors, this could be explained by the experimental groups' frequent production of modified output in reaction to feedback, which may enhance learners' noticing, hypothesis testing, and automatizing of correct forms.

Previous studies have produced evidence for the link between (modified) output and noticing. For example, Izumi et al. (1999) investigated the effect of output on noticing and subsequent acquisition of grammar forms. The experimental group received written input as well as the opportunity to produce output, while the control group only received input and comprehension questions. The intervention included two phases with two input-output cycles each and was followed by noticing and acquisition assessment after each cycle. Noticing was assessed by asking participants to underline relevant forms in the input. Language gains were measured by post-tests on the targeted grammatical form. The results revealed that noticing increased equally for both the control and the treatment group. The authors suggested that increased

noticing in the control group was stimulated by the comprehension questions which drew the learners' attention to the target forms during the subsequent input presentation, whereas the treatment group learners' noticing was enhanced by producing output. This suggests that output production does in fact promote noticing, although the unique effects of output on noticing were not confirmed. Furthermore, learners from the treatment group incorporated the target forms in their production during the output cycle during phase 1. However, in the phase 1 post-test the learners showed no improvement in their use of the target form. The authors suggest that the learners could not process the target form sufficiently due to the task's high cognitive demands. For phase 2, learners from the treatment group showed significant improvement on the post-test whereas the control group did not. Again, the authors explain this difference with different cognitive demands in the two groups. To conclude, the findings of this study suggest that in order for output to enhance noticing and lead to acquisition, the task may not place heavy cognitive demands on the learners.

Corroborating the claim on output and noticing, Egi (2010) found that learners' noticing of the gap between their production and the target language was significantly greater when they produced repair rather than uptake without repair following feedback. According to the author, this suggests that repair could be more closely linked to the cognitive process of noticing the gap than uptake and might explain why repair is more strongly associated with L2 development than uptake in general.

Some studies investigated whether monitoring and internal feedback as opposed to external feedback may equally lead to modified output and (self-)repair. For example, Shehadeh (2001) examined the role of monitoring – referred to as “self-initiation” (p. 442) – in prompting modified output. The participants, 35 adult EFL learners, interacted orally with peers or native speaker pairs on three different communication tasks. The results demonstrated that both feedback as well as monitoring triggered modified output in a large majority of the cases. It was further found that monitoring led to modified output to a higher degree than other-initiation did in four out of five of the interactional contexts. In addition, in almost all instances (93%), the modifications resulted in successful self-repair. The author concluded that monitoring may lead to greater opportunities for modified output than external

feedback and thus, besides feedback, constitutes an important feature for FL learning. Sato (2012) and Sato and Takatsuka (2016) similarly demonstrated that in learner-native speaker interactions self-initiated self-repair was frequent and in general successful.

Investigating the direct effect of output on L2 development, McDonough (2005) showed how modified output following negative feedback affects the acquisition of L2 forms, operationalized here as question formation. The participants engaged in interactional tasks with native speakers under four different conditions over the course of eight weeks: 1) Native speakers enhanced the salience of non-target-like questions through repetition and clarification requests such as “huh?” or “sorry?”; 2) native speakers requested clarification without identifying the problematic structure; 3) native speakers repeated the problematic features without giving the learners time to modify their output; 4) native speakers provided no feedback when learners produced non-target-like questions. The impact of feedback and modified output on question development was measured in oral production pre- and post-tests. Modified output involving question forms at least one stage above the current stage were considered for analysis. The results revealed that modified output was the only significant predictor of question development. In addition, feedback in the form of clarification requests positively correlated with question development and modified output. The findings thus suggest that modified output predicts question development, while clarification requests may play an indirect role by facilitating modified output.

Interactional Features

In the context of FL and L2 interaction, interactional features describe those features of the interaction that involve a “communicative aspect” (Ellis, 1991, p. 180) such as negotiation of meaning, feedback, and alignment. Some of these features have already been touched on in the sections above in terms of their role with respect to input and output but will be explained in detail below.

Negotiation of Meaning

As explained above (see section Historical Origins), negotiation of meaning refers to the process in which either interlocutor, i.e., the speaker or the hearer, signals a communicative problem and both engage in interactional work in order to achieve mutual understanding. The model by Varonis and Gass (1985) is the most widely used model of negotiation of meaning sequences (Smith, 2003, p. 39). It conceptualizes negotiation of meaning sequences as consisting of four components: trigger, indicator, response, and reaction to response, as in the example taken from Varonis and Gass (1985, p. 77):

S: My father now is retire. [trigger]
 J: retire? [indicator]
 S: yes [response]
 J: oh yeah [reaction to response]

As seen in the example, the trigger refers to the utterance or part of an utterance such as a specific lexical item or grammatical structure that leads to some degree of non-understanding on the part of the interlocutor. The trigger is followed by a signaling of non-understanding, i.e., the indicator, uttered by either interlocutor. If indicated by the hearer, its function is to signal that the input was not understood, indicators uttered by the speaker signal that they are uncertain if they have been understood, as in the case of foreigner talk (Ellis, 2008). The indicator may be realized in different ways. Long (1983) observed three key strategies which have since been the most commonly investigated indicator types in SLA research studies (Loewen & Sato, 2018): confirmation checks (used by the hearer to ensure that they have understood the speaker correctly), clarification requests (to elicit from the speaker additional information about the meaning of the problematic utterance), and comprehension checks (used by speakers to prevent an anticipated communication breakdown). The non-understanding is acknowledged in the response. The literature describes different ways in which responses are realized. Long (1983), for example, identified strategies such as reformulations, narrowing down of the topic, repetitions, and or-choice questions. The response may be followed by an (optional) reaction to the response which serves as a signal that learners are ready to resume their conversation and

typically taking the form of an explicit statement of understanding such as “ok” or “I understand”.

The cognitive-interactionist approach associates negotiation of meaning with a number of benefits for FL and L2 learning. As mentioned above (see section Input), a first benefit is that it can help to make input comprehensible to the learner. Through negotiation of meaning, comprehensibility may furthermore be achieved without reducing the input’s linguistic complexity thanks to overall conversational adjustments (Long, 1996; Mayo & Soler, 2013).

When initiated by the interlocutor, negotiation of meaning may also function as a type of negative feedback as the interlocutor’s indication of non-understanding signals to the learner that their production was in some way problematic. The benefits of feedback will be further detailed below (see section Feedback).

As a more general effect, negotiation of meaning may increase the learner’s focus on form, i.e., the linguistic elements of the interaction, while otherwise engaged in a primarily meaning-centered communicative task (Long, 1991). Focus on form, in turn, may allow learners to briefly pause their focus on meaning and attend to linguistic items that might otherwise go unnoticed. Since focus on form is always embedded in a meaning-based activity, learners’ attention may be particularly drawn to the form-meaning relationships, i.e., the relationship between linguistic forms and the meaning they convey (Mayo & Soler, 2013).

The claims on the different benefits of negotiation of meaning have been empirically investigated in various studies. The effect of negotiation of meaning as one type of input modifications on input comprehensibility was already reviewed above (see section Input). As summarized above, the findings from the studies point to a beneficial effect of interactionally modified input, i.e., negotiated input, on comprehensibility as well as an advantage over pre-modified input.

Further studies have provided evidence for the positive impact of negotiation of meaning on noticing. For example, Pica et al. (1996) investigated modified input, modified output, and feedback in negotiation of meaning sequences in L2 interaction. Each of the participants was randomly assigned to another learner or a native speaker

to interact on two communicative tasks. One of the findings was that in both groups the learners' signaling of non-understanding predominantly consisted of segmented parts of the interlocutor's previous utterance (e.g., "the house has two maybe two stone steps" – "two stone steps"?). According to the authors this may indicate that during negotiation of meaning the saliency of specific linguistic features in the input increases for the learner.

Bitchener (2004) showed that negotiation of meaning also heightens the learners' noticing of problematic items in their own production. In the study, L2 learners interacted in peer dyads on two different communication tasks. They performed the same tasks three times: at the initial session, the second session one week later, and the third session twelve weeks later. The learners' retention of negotiated vocabulary, pronunciation, and grammar items was measured in tailor-made tests three days after the third session to assess the impact of the negotiation on acquisition. Analysis of the negotiation of meaning sequences demonstrated that learners noticed the gap between their own production that triggered negotiation of meaning and, in a majority of instances, they immediately modified their problematic utterance after receiving a signal of non-understanding. It was furthermore found that a reasonably high proportion of the modified items were retained after one week as well as after twelve weeks, suggesting that L2 learning facilitated by negotiation of meaning can be retained in the long-term. The positive impact of modified output following negotiation of meaning on acquisition was also shown by McDonough (2005), which was described in detail above (see section Output).

Most experimental research on negotiation of meaning, however, has aimed to establish a direct link with acquisition rather than finding indirect evidence by exploring its hypothesized benefits in providing comprehensible input, enhancing noticing, and opportunities to produce modified output. For example, the study by Ellis et al. (1994) mentioned above (see section Input) investigated the role of negotiation input on the acquisition of vocabulary. The paper reported on two studies, the Saitama Study and the Tokyo Study, using the same design, where participants were assigned to the control group (unmodified input), a group exposed to pre-modified input, or a group that negotiated input in interaction. The researchers identified 18 unknown lexical items in a translation pretest completed by the

participants, which were subsequently seeded into the treatment task. When performing the task, the participants listened to directions that varied according to the three groups. The control group received unmodified directions, whereas the two treatment groups were given pre-modified input (more repetitions of the target-items, slower speech) or had a chance to request clarification from the teacher, respectively. Following the directions, the participants placed objects on a picture. The participants' comprehension was measured by means of their success in placing the objects correctly on the picture. In both studies, the interaction group achieved significantly higher comprehension than the control and pre-modified groups. The interaction group also outperformed the other groups in two post-tests (two days as well as one month after treatment) and a follow-up test (2.5 months after treatment) measuring the acquisition of the target items. However, in the Saitama study, no comparison to the baseline/control group was possible, and in the Tokyo study, the interaction group outscored the pre-modified group only in the post-tests but not the follow-up test, pointing towards benefits for vocabulary only in the short term.

A study by De la Fuente (2002) additionally highlights the role of output for the positive effect of negotiation of meaning on acquisition. In the study, 32 university Spanish learners were assigned to one of three groups: 1) non-negotiated pre-modified input; 2) negotiated input without output; 3) negotiated input plus output. Each group interacted with a native speaker of Spanish on two tasks. In both tasks, the learners were given oral instructions by the native speaker to locate and identify target words in a room. Group 1 (non-negotiated) was given the instructions at a slow speed, group 2 (negotiated input, no output) was allowed to negotiate the meaning of the target words, and in group 3 (negotiated input plus output) learner and native speaker reversed their roles in the second half of the task, allowing the learner to produce (modified) output. Each groups' comprehension was measured by their ability to follow the instructions accurately during the task. Additionally, a post-test measured the impact on vocabulary acquisition immediately after the treatment, one week later, and three weeks later. The results showed that comprehension was significantly higher for learners from the two negotiation groups compared to the non-negotiation group. With regard to receptive vocabulary acquisition, the only significant group difference was between the negotiation plus output group and the non-negotiation group with

the former outperforming the latter. The opportunity to produce output also played a decisive role on productive vocabulary acquisition: the negotiation plus output group outperformed both the group without negotiation and the group with negotiation but no output. However, there was no significant difference between the group with no negotiation and the group with negotiation. In sum, the study demonstrated that negotiation improved comprehension, that negotiation (but not additional output) led to higher receptive acquisition, and that productive acquisition relies on the production output in addition to negotiation.

Focusing on the feedback function of negotiation of meaning and its impact on acquisition, several studies have shown a positive effect. For example, a small-scale study by Nobuyoshi and Ellis (1993) suggested that the negotiation of meaning indicator type clarification request may contribute to acquisition by leading learners to produce more accurate output. In the study, six adult English learners interacted on two picture jigsaw communication tasks with their teacher. The three participants from the experimental group received clarification requests each time they produced an incorrect target item, i.e., past tense, whereas the three participants from the control group received only general clarification requests and not specifically on incorrect past tenses, i.e., when the teacher genuinely failed to understand an utterance. The findings revealed that two of the learners from the experimental group reformulated and corrected their erroneous utterances after receiving clarification requests, while the third learner corrected hardly any of their errors. In addition, the two learners produced substantially more accurate past tense verbs during a second administration of the same task where the teacher made no requests to reformulate incorrect past tense verb forms. The third learner made no accuracy gains during the second administration. In the control group, there were no corrected reformulations or more accurate production during the second administration. The findings suggest that learners may benefit from clarification requests on target items in that it leads them to produce more accurate output and make acquisitional gains.

Other studies investigated the relative effect of negotiation of meaning among other feedback types. In their meta-analysis on interaction research, Mackey and Goo (2007) combined the results of the acquisitional effects on the three most studied feedback types, i.e., recasts, negotiation of meaning, and metalinguistic feedback. They

found a medium mean effect size on immediate post-tests for negotiation ($d = 0.52$), compared to a very large effect for recasts ($d = 0.96$) and a small to medium effect for metalinguistic feedback ($d = 0.47$). However, the authors indicate that due to the limited number of available studies any definite conclusions would be premature.

To sum up the findings of the research on negotiation of meaning, they have provided evidence for the benefits of negotiation on input comprehensibility and increased noticing of items in the input as well as the beneficial role of negotiation of meaning for the production of modified output and associated acquisition of the negotiated items. There is also direct evidence for the benefits of negotiation of meaning on the acquisition of L2 grammar and lexical items, which has further been shown to be enhanced by the production of the target items through modified output.

Feedback

Another key component and argued benefit of interaction is feedback, which, in the context of interaction refers to the interlocutor's response to an error in the learner's utterances, providing learners with information about the success or lack thereof of their output (Gass & Mackey, 2007; Loewen, 2011; Mayo & Soler, 2013). While feedback can be positive or negative, cognitive-interactionist research typically focuses on negative feedback (Mayo & Soler, 2013), which is also referred to as *corrective* feedback. Corrective feedback may target phonological, lexical, grammatical, or pragmatic aspects of learner speech (Lyster et al., 2013). While corrective feedback is most common in conversations between learner and teacher, it may also occur in learner-native speaker or learner-learner interaction (M. Sato & Lyster, 2012).

There are several ways in which corrective feedback is provided during interaction: one broad distinction is made between implicit and explicit feedback. When implicit feedback is provided, the focus of the conversation usually remains on meaning and there is no overt indication of an error (Ellis et al., 2006). Implicit feedback can for example take the form of negotiation of meaning (see section Negotiation of Meaning), more precisely through confirmation checks (e.g., "My father is retire." – "Retire?") or clarification requests (e.g., "I'm sorry?") by the interlocutor. Implicit feedback can also take the form of repetitions or elicitations/prompts (e.g.,

“He kiss her.” – “He ...?”), encouraging learners to self-correct without providing them with the correct linguistic forms. Another form of implicit feedback are recasts, i.e., more target-like repetitions/reformulations of a learner’s incorrect utterance that maintains its original meaning (Loewen, 2011) (e.g., “How many is it?” – “How much is it? It depends.”). As this example illustrates, recasts may in some cases be part of a negotiation sequence and take the function of a confirmation check (Loewen & Philp, 2006; Lyster et al., 2013).

Explicit feedback involves the overt indication of an error, as in corrections (e.g., “It’s not *were*, it’s *was*”) and metalinguistic comments/clues (e.g., “You need singular”). Explicit types of feedback are typical of learner-teacher interaction and do not usually occur in learner-native speaker interaction (A. Mackey, 2007).

As mentioned above, feedback can further be distinguished as providing positive or negative evidence to the learner. For example, recasts are instances of feedback that provide learners with positive evidence, revealing to them what is accurate in the target language. Negative evidence, for example, provided through negotiation of meaning or explicit corrective feedback, has the function of informing the learners on what is not accurate.

As can be seen in these examples and definitions, feedback can either be output-prompting, i.e., elicit the correct form from the learner, or input-providing, i.e., provide the correct form to the learner.

As already explained above (see section Output), feedback may be followed by learner uptake, i.e., the learner’s utterance following and containing a reaction to the feedback (Egi, 2010), which may more or less successfully modify the erroneous form as *repair*, *partial repair*, or *no repair/needs repair*.

Corrective feedback is argued to provide learners with a range of benefits that vary depending on the specific feedback type. First of all, negative feedback provides learners with negative evidence on their erroneous production and allows them to notice mismatches between their own erroneous utterance and the target-like form. Input-providing feedback may additionally provide learners with positive evidence (Loewen, 2011). It has been suggested that this is particularly necessary for linguistic items that might otherwise be non-salient in the input, for example because their

erroneous use does not hinder communication (White, 1991). Based on this evidence learners may test and revise their hypotheses about the FL/L2 system (Loewen & Sato, 2018; Sheen & Ellis, 2013).

Feedback may furthermore prompt learners to modify their output, which plays a facilitative role in L2 acquisition, as shown above (Output). In the case of output-prompting feedback, learners are additionally required to self-correct their utterances (Loewen, 2011).

One of the major questions of research on corrective feedback in interaction has been what type of feedback and which feedback characteristics might be most beneficial for acquisition. This question has been approached either indirectly by investigating what type of corrective feedback is best noticed by the learner, or directly, by measuring the effect on FL/L2 development.

Studies taking the indirect approach by examining the relationship between different types and characteristics of feedback and the degree to which they are noticed by learners, have typically focused on learner uptake as evidence for noticing. Lyster and Ranta (1997) for example observed teacher feedback in six French immersion classrooms in Canada and the degree to which different feedback types led to uptake and repair. The authors identified six types of feedback with recasts being the most frequently used feedback type (55%), followed by elicitation (14%), clarification request (11%), metalinguistic feedback (8%), explicit correction (7%), and repetition (5%). However, recasts produced the least amount of uptake (31% uptake), which the authors interpreted as a possible indication that learners do not necessarily perceive them as feedback but rather repetition. Explicit correction was equally unsuccessful (50% uptake), while elicitation (100% uptake), clarification requests (88% uptake), metalinguistic feedback (86% uptake), and repetition (78% uptake) were the most successful feedback techniques in eliciting uptake, which according to the authors, might suggest that these feedback types are less ambiguously perceived as feedback by the learners. The authors also point out that not all feedback types allow for student-generated repair, i.e., self-repair, and thus to engage the learners actively in the negotiation of form. They found that elicitation and metalinguistic feedback led to most self-repair (45% and 46%, respectively, of total repair), followed by clarification

requests and repetition (27% and 31%, respectively, of total repair). Recasts and corrective feedback by providing the correct form did not elicit any self-repair. The study thus suggests that the degree to which learners uptake, i.e., notice, and repair the targeted features depends on the feedback type.

Ellis, Basturkmen, and Loewen (2001) similarly investigated learner uptake in different types of feedback. Analyzing classroom talk, the authors identified three types of feedback: teacher-initiated corrective feedback and two preemptive rather than reactive feedback types, i.e., student-initiated focus on form and teacher-initiated focus on form. The results showed that feedback type influenced the amount of uptake, with uptake being highest in student-initiated focus on form, intermediate in teacher-initiated corrective feedback, and lowest in teacher-initiated focus on form. According to the authors this indicates that students might be more likely to attend to form when they themselves perceive the problem as important. Furthermore, complex episodes where the first attempt at resolving a linguistic issue failed, resulted in more uptake and more successful uptake than simple episodes. Uptake was also higher for episodes involving negotiation of meaning, according to the authors possibly because they dealt quite explicitly with lexical problems. Similarly, direct feedback providing explicit information led to more uptake than recasts, an indirect feedback type. According to the authors, this might indicate that learners are more likely to notice linguistic features when they are treated explicitly. The authors also point out that the relatively high rate of uptake (71.6%) and repair (76.3%) for recasts stands in contrast to the findings in Lyster and Ranta (1997). In sum, the study demonstrates that not only feedback type but the specific characteristics of the feedback impacts on the degree to which the target forms are noticed, with higher noticing in linguistic features that learners themselves perceived as important, when they had the opportunity to negotiate them extensively, and when they were treated explicitly rather than implicitly.

With respect to recasts, Sheen (2004) investigated the specific characteristics influencing their effectiveness in eliciting uptake and repair further. Exploring the occurrence of feedback in four different contexts (1. French Immersion; 2. Canada ESL; 3. New Zealand ESL; 4. Korean EFL) the author found that recasts are the most frequently used type of feedback, in line with previous research (e.g., Lyster & Ranta,

1997). Analysis of the relationship between feedback types and rate of uptake and repair indicated that the extent of uptake and repair depended on factors such as the students' age, L2/FL proficiency, and educational background may also play a role. For instance, both uptake and repair, and thus noticing of feedback, were more prevalent in the contexts with educated adults than those settings with children and less educated adults. With respect to recasts, noticing was impacted by their specific nature, in particular, their salience and explicitness. Thus, recasts were more likely to be followed by uptake the more salient they were, e.g., by isolating the target feature. Additionally, context variables such as pedagogical focus as well as age, proficiency, and educational background of the learners played a role in the effectiveness of recasts.

Subsequent studies confirmed the complex relationship between interactional feedback and uptake. For instance, Loewen and Philp (2006) similarly found that in their comparison of the three feedback types elicitations, informs (i.e., an explicit explanation about the problematic form), and recasts, the degree of uptake depended on feedback type. However, in their study, elicitations led to the highest rate in successful uptake (i.e., uptake incorporating the feedback), followed by recasts, and last, informs. The authors explained the relatively high uptake rate of recasts compared to previous studies with the fact that recasts in their study was characterized by low implicitness, e.g., by being stressed by the speaker or by being isolated. They also observed a high overall uptake rate across all feedback types, which they linked to the studied setting involving "highly interactive adult language classrooms" (Loewen & Philp, 2006, p. 549).

Lyster and Mori (2006) also found varied uptake and repair rates following recasts and prompts, respectively, in Japanese and Canadian immersion classes: Whereas in the Japanese classes recasts were more conducive in leading to uptake and repair, prompts were more successful in the Canadian classes. The authors suggested that in meaning-oriented settings such as the Canadian classrooms, prompts were more effective as they directed the learners' attention to form, whereas in form-oriented settings, as in the Japanese classrooms, learners are sensitive to implicit feedback and can benefit from recasts.

More recent research has additionally investigated the relationship between uptake and noticing, on the one hand, and uptake and acquisition, on the other hand.

Concerning the relationship of uptake and noticing, it was found that while uptake can be seen as an indicator of learner noticing (Egi, 2010), the absence of uptake does not necessarily mean that no noticing had occurred (Ellis & Sheen, 2006; Goo, 2020; Li & Vuono, 2019). Thus, Yoshida (2010), for example, observed that learners reported more noticing of recast feedback during stimulated recall than uptake was measured in the interaction scripts.

With regard to the role of uptake for FL and L2 development, findings highlight the function of uptake in producing modified output, especially in response to output-prompting rather than input-providing feedback. For example, the study mentioned above by Loewen and Philp (2006) found that successful uptake, i.e., uptake containing the repaired form, predicted learners' scores in the post-test. However, this was only for the output-prompting feedback types elicitations and prompts, but not for the input-providing feedback type recasts. Similarly, Nassaji (2011) studied the differential effect of two types of uptake (repetition or incorporation of the target form) on the learners' post-test scores. The author observed that incorporation was more beneficial than repetition. The findings from both studies suggest that uptake that requires learners to draw on their own knowledge, for example when producing the target-like form on their own in response to a prompt or when incorporating it in a new linguistic context, might be more effective for acquisition than when the target-like form is merely repeated by the learner, for example in response to a recast.

Another body of research has directly investigated the effectiveness of different feedback types for acquisition. In most studies, acquisitional gains have been operationalized as increased accuracy in the use of the targeted linguistic item when performing a production test, grammaticality judgment test, or tailor-made test (Loewen, 2011). Moreover, much of the research has focused on recasts, probably because they constitute the most frequently used type of feedback used by teachers in language classrooms (Brown, 2014).

The results on the relative effectiveness of different feedback types, and recasts in particular, have been mixed. Some studies found recasts to be less beneficial for the acquisition of linguistic structures than more explicit types of feedback such as metalinguistic feedback (Ellis et al., 2006) or prompts (Yang & Lyster, 2010). Others found no difference for the type of feedback, e.g., McDonough (2007) who compared

recasts and clarification requests, Goo (2012) on the effects of recasts and metalinguistic feedback on the acquisition of questions using the English *that*-trace filter (note that in Goo's study, learners were blocked from producing modified output in both treatment groups), or Loewen and Nabei (2007) who compared the effects of recasts, clarification requests, and metalinguistic feedback on the emergence of simple past activity verbs.

It has been suggested that these heterogeneous findings on the effectiveness of various types of feedback are due to the complex impact of feedback on learning and its interaction with a number of individual as well as context variables. Factors that have been found to mediate the effect of feedback include individual differences such as proficiency (e.g., S. Li, 2013, 2014), working memory (e.g., Goo, 2012; S. Li, 2013; A. Mackey & Sachs, 2012), and anxiety (e.g., S. Li, 2013; Sheen, 2008) as well as contextual variables including the type of linguistic target (e.g., S. Li, 2014; Yang & Lyster, 2010; Yilmaz, 2016), the specific characteristics of the feedback type (e.g., Ellis & Sheen, 2006), the interlocutor's experience and aptitude in giving feedback (Junqueira & Kim, 2013), as well as the instructional setting (Lyster & Mori, 2006). With respect to the effectiveness of recasts, in particular, it is important to note that, as pointed out by Goo and Mackey (2013), despite the relative disadvantage of recasts compared to other types of feedback in some studies, there is substantial evidence for the beneficial effects of feedback in general, including recasts for acquisition brought forward by all studies.

To illustrate these points, two studies on different feedback types will be discussed in more detail. First, the study by Ammar and Spada (2006) compared the effects of recasts and prompts on the acquisition of possessive determiners *his* and *her* in French-speaking ESL learners. Both the recast and prompts treatment groups significantly outperformed the control group, who received no feedback, in the written and oral post-tests. The results also revealed that, overall, prompts were more beneficial than recasts. However, the effectiveness of recasts depended on the learners' proficiency: learners who demonstrated more accurate use of the target structure in the pretest benefitted equally from both types of feedback, while those with less accurate use benefitted significantly more from prompts than recasts. In sum, this study demonstrates that learners in general benefit from feedback and that, while one

specific feedback type (prompts) has been found to be more beneficial than the other (recasts), this was only true for lower-proficiency learners.

Lyster and Izquierdo (2009) reported that participants made progress in their use of grammatical gender in French, regardless of whether they received feedback in the form of recasts or prompts or not. The authors attribute several reasons to the fact that the results contrast with those of previous studies: The recasts used in the study were relatively explicit (they consisted of isolated erroneous noun phrases) and the study took place under controlled conditions in a laboratory as opposed to a classroom setting. Moreover, learners receiving recasts benefited from repeated exposure to positive exemplars in addition to being able to infer negative evidence, whereas learners of the prompts group benefited from repeated exposure to negative evidence and opportunities to produce modified output. Therefore, as this study illustrates, the relative advantage or disadvantage of a specific feedback type (here, recasts) is mediated by several factors such as the specific characteristics of the feedback type and the setting.

In addition to the evidence brought forward by individual studies, several meta-analyses have confirmed the observations about the general effectiveness of feedback as well as the relative effectiveness of specific feedback types and characteristics. Li (2010), for example, meta-analyzed 33 classroom and laboratory studies, finding a medium overall effect on L2 learning ($d = 0.61$). Furthermore, the study found larger immediate effects for explicit (i.e., metalinguistic feedback and explicit correction) than for implicit feedback (i.e., recasts, clarification requests, elicitation, and repetition). With regard to the long-term effects, however, due to the small sample size being tentative, the results indicated an advantage of implicit over explicit feedback.

The meta-analysis by Lyster and Saito (2010) only included classroom studies. The authors similarly found a medium to large overall effect for feedback. With respect to the different feedback types, all investigated feedback types, i.e., prompts (here used as an umbrella term for clarification requests, repetition, elicitation, and metalinguistic clues), recasts, and explicit corrections were all found to be significantly effective for classroom learning with no significant decline of the long-term effect. Comparing the effectiveness of the different feedback types, the authors found significantly larger

effects for prompts than for recasts and explicit corrections. Since both prompts and explicit corrections are explicit types of feedback, this finding does not contribute to the question whether explicit or implicit feedback is more advantageous. It does, however, suggest that output-prompting feedback (i.e., prompts) might have an additional benefit to classroom learners compared to input-providing feedback (i.e., recasts).

To sum up the current state of research on feedback, research focusing on the relationship of feedback and output has found that uptake can be seen as evidence for noticing. However, a high uptake rate does not correspond one-to-one with high FL/L2 gains: On the one hand, noticing may occur in the absence of uptake. On the other hand, the beneficial effect of uptake depends on the quality of the uptake, with uptake having the greatest benefits when learners are pushed to produce the target-like form by drawing on their own resources, i.e., when the interlocutor does not provide the target-like form, through incorporation in a new linguistic context. Furthermore, the degree of noticing varies, as evidenced by uptake, depending on feedback type and characteristics as well as on factors relating to the learner and setting. Thus, noticing tends to be greater for feedback that is considered relevant by the learner and that is explicit and salient, however, implicit types (e.g., recasts) might work for adult learners, higher-proficiency learners, and more educated learners, and in form-focused classrooms.

Research investigating the direct impact of feedback on FL and L2 development has provided robust evidence for the overall beneficial effect of feedback on learning across the different feedback types as well as for all individual feedback types. The findings on the relative effectiveness of the different feedback types have been varied, indicating that the affordances of the different types depend on the learner and setting. Similarly, to the findings on feedback and uptake above, the evidence suggests that especially low-proficiency learners might benefit from explicit feedback and that output-prompting feedback is generally advantageous.

Alignment

The concept of alignment has recently entered cognitive-interactionist research from psycholinguistic first language research (Pickering & Garrod, 2004) to explore its occurrence and relation to FL and L2 learning through interaction (Y. Kim, 2017). In psycholinguistics, the term *alignment* is used to describe the phenomenon of interlocutors reusing each other's language patterns in dialogue. According to Pickering and Garrod (2004) alignment follows a largely automatic process of interlocutors converging on the use of linguistic elements and structures leading to the alignment of their situation models, i.e., the mental representations of the situation that is being discussed as it relates to space, time, and intentionality (Zwaan and Radvansky, 1998, as quoted in Pickering & Garrod, 2004, p. 172). They further claim that alignment is driven by priming, the cognitive activation of a particular mental representation through an utterance (= prime) making it more likely that the speaker will subsequently produce in their speech an utterance that uses that representation (= target) (Pickering & Garrod, 2004).

In the FL and L2 literature the term alignment has sometimes been used interchangeably with priming (Kim et al., 2020). A further terminological distinction needs to be drawn to entrainment (Pickering & Ferreira, 2008), the actual manifestation of alignment in form of a repeated speech feature in dialogue (Michel & Cappellini, 2019). In the present study, the term *alignment* will be used in line with previous FL research to refer to the repetition of a linguistic feature used in previous discourse as the result of priming (Michel & Cappellini, 2019).

In FL and L2 interaction, it is assumed that the processes underlying alignment are not purely automatic but may also be controlled and strategic (Costa et al., 2008). Firstly, as Costa et al. (2008) suggest, automatic alignment might be reduced in FL and L2 speakers since they are less proficient in the language compared to native speakers. Thus, FL and L2 speakers might fail to align to their interlocutor if they are less familiar with the linguistic feature that was used compared to a more familiar and automatized alternative. Similarly, their automatic alignment might be impaired by the increased processing load when under pressure to speak faster than they are able to. Secondly, FL and L2 speakers may deliberately avoid certain linguistic features that

they cannot control sufficiently, e.g., to avoid inaccuracy. In a similar way, FL and L2 learners may tend to align to linguistic features that are similar to those in their first language (L1) but avoid alignment of features that are less similar. Conversely, alignment in FL and L2 interaction may increase when learners choose to align with their interlocutor in order to facilitate learning. When interacting with a native speaker, for example, learners may deliberately repeat the native speaker's well-formed examples with the goal of producing and practicing target-like output and testing hypotheses, as well as to continue the dialogue to ensure the continued flow of input. Native speakers, on the other hand, may deliberately align to the learner's lexical and syntactic choices to accommodate them, e.g., when repeating high frequency vocabulary and simple syntactic structures in foreigner talk, or to signal understanding. In FL/L2 peer interaction, automatic alignment is more likely when the interlocutors' first languages are closely related. However, since in FL/L2 peer interaction the interlocutors need to constantly assess the other's proficiency level and evaluate if their linguistic choices will be understood, automatic alignment in FL/L2 peer interaction can be expected to be inhibited as the interlocutors' assessment occurs at the cost of attentional resources.

While alignment may involve all aspects of converging behavior ranging from the reuse of the same vocabulary or syntactic structures to the mimicking of gestures and body posture (Michel & Cappellini, 2019), the focus of alignment research in the field of FL and L2 research is on linguistic alignment, i.e., the phenomenon in which the speakers reuse lexical material, grammatical structures, or phonological features that they have previously experienced in the conversation (Kim et al., 2020).

Linguistic alignment is based on the premise that speakers will produce lexical material or a grammatical structure that they have previously been exposed to although an alternative structure could be used to express similar meaning (e.g., passive instead of active structure) (Y. Kim, 2017). The reuse of lexical material is generally referred to as lexical alignment. Syntactic or structural alignment describes the reuse of a grammatical structure such as the use of a passive construction (e.g., "The chairs can't be moved.") after being exposed to that structure previously in the interaction (e.g., "A confirmation email will be sent to you.").

Alignment in FL and L2 research has distinguished between *self-alignment* or *within-speaker alignment*, i.e., the speaker's alignment to their own previous production, and *other-alignment* or *between-speaker alignment*, i.e., alignment to the interlocutor's production (Dao et al., 2018; Michel & Cappellini, 2019).

Note that instances where the interlocutor aligns to a learner's erroneous utterance, thereby correcting the error and producing a target-like reformulation of the utterance, are referred to as recasts (Costa et al., 2008), otherwise considered a type of feedback (see section Feedback).

From a cognitive-interactionist perspective, one theoretical benefit of linguistic alignment for language development relates to the modeling of target-like input and opportunities to produce output. More precisely, it has been argued that alignment may be facilitative in eliciting specific linguistic features in learner production (Loewen & Sato, 2018). As noted above (see section Output), the repeated production of linguistic features in turn might contribute to their automatization (McDonough & Mackey, 2008). However, it has also been pointed out that learners need to already have stored the respective knowledge representation and that the features cannot be entirely new to the learner (McDonough & Mackey, 2008; Trofimovich et al., 2014).

As an additional possible benefit, alignment may particularly promote the development of more abstract syntactic representations (McDonough & Mackey, 2008). This argument is based on the concept of item-based learning in L1 acquisition (Tomasello, 2000) which suggests that syntactic development proceeds from formulaic expressions where syntactic representations are associated with specific lexical items to more abstract representations through repeated exposure to the structure. In this view, alignment might encourage learners to produce a given structure in a variety of lexical contexts, thus ultimately leading them to form more abstract representations of new syntactic knowledge (Kim et al., 2020; McDonough & Mackey, 2008).

Research on FL and L2 learner alignment conducted over the past decade has shown that priming does occur in FL and L2 interaction and may successfully lead to a higher or more accurate production rate of a variety of target structures such as, for instance, prepositional datives (McDonough, 2006), passive voice (Kim & McDonough,

2008), question formation (McDonough & Chaikitmongkol, 2011; McDonough & Mackey, 2008), pronunciation (Trofimovich et al., 2014), or verb inflection (Marsden et al., 2013). In many of these studies, the interlocutor used a script to provide primes triggering the use of the target structure by the learner. Acquisitional gains have been measured through pre- and post-tests.

McDonough and Mackey (2008), for example, investigated the occurrence and effect of structural alignment on the development of English question formation. In their study, research assistants engaged in communicative tasks with ESL learners using scripted question forms that were one developmental stage higher than the learners' current knowledge of question formation. The evidence showed that participants produced more developmentally advanced questions after the advanced questions primes than after same or lower-level questions. Comparing the results of pretest, post-test, and delayed post-test four weeks later, learners with high levels of syntactical priming during the tasks tended to produce question forms a stage higher than the learners' knowledge of question formation.

Recently, researchers have shifted their focus from lab-based alignment studies to alignment in authentically occurring learner interaction in classrooms, with the goal of exploring its role and potential for the classroom learning context.

For example, Trofimovich et al. (2014) explored the alignment of stress patterns during peer interaction in an L2 classroom. Participants collaborated in information-exchange tasks, seeded with three- and four-syllable words with stress on the second syllable. Transcripts of the peer dialogues were analyzed for alignment which was operationalized as the occurrence of an accurate stress pattern following an accurate prime. The results revealed that across all four tasks the learners produced a higher proportion of accurate stress patterns when preceded by an accurate stress pattern in their interlocutor's turn than they did when the interlocutor had not produced an accurate stress pattern. Alignment was observed in the pronunciation of the same words as well as of different words with the same stress pattern, with the majority of alignment occurring across different words. However, it was also found that, albeit rarely, learners aligned to their interlocutor's inaccurate stress patterns, indicating that alignment in L2 peer interaction may also reinforce learner errors. In conclusion, the authors point out that while the study showed that pronunciation alignment occurs in

classroom peer interaction providing learners with ample opportunity to repeatedly hear and produce the target stress patterns, its potential learning benefit would need to be investigated in further effectiveness studies.

McDonough, Trofimovich, and Neumann (2015) in a similar study on structural alignment created activities to elicit written and oral production of target structures in English for academic purpose classes. While the treatment group received activities seeded with the target structures (primes), the control group received activities without primes. The study tested alignment for three syntactic structures: relative clauses, adverbial clauses, and passives. The results indicated that the treatment group produced a significantly greater proportion of target structures when primed to do so than in unprimed contexts, suggesting that structural alignment does occur in authentic classroom activities. It was furthermore found that the treatment group produced more target structures than the control group; however, only in oral production and only with regard to relative clauses and adverbial clauses, but not for passives. The authors suggested that passives were elicited naturally in the activities whose meaning could often not be expressed by an alternative construction, which reduced the differences in alignment between the two groups. However, relative and adverbial clauses are considered difficult and infrequent structures and were thus much less likely to be produced by learners in the unprimed activities. From this observation, the authors conclude that structural alignment is particularly useful when aiming to elicit difficult or infrequent structures.

More recently, Dao et al. (2018) explored structural alignment in less scripted communicative tasks without predefined target structures typically delivered by a researcher acting as interlocutor. The participants, non-native English students at a Canadian university, interacted in dyads on two communicative tasks. Structural alignment was operationalized not as the immediate repetition of a target structure from the interlocutor's speech as in the cited studies above. Instead, structural alignment was operationalized following Pickering and Garrod as instances where one speaker reuses a structure within six turns after the same structure was produced by either the interlocutor (between-speaker alignment) or the same speaker (within-speaker alignment). Instances of alignment were coded in the conversation transcripts for type (e.g., existential *there*, *that* complement, relative clause) as well as accuracy

and alignment between or within speakers. For each dyad, the authors subsequently calculated a score of prime-target sequences as well as a score for no prime-target sequences for those instances where a speaker produced a target without being primed within the preceding six turns. The results revealed significantly more instances of prime-target sequences than no prime-target sequences, with within-speaker alignment clearly being the dominant type of alignment compared to between-speaker alignment. While the extent of alignment was comparable in both tasks, the aligned structures varied (e.g., the picture task involved alignment of *that* complement, the map task yielded *go + PP* as the most frequent structure). In addition, alignment was linked to improved production accuracy, in that learners tended to produce accurate structures at a higher rate when the accurate structure was used within the six preceding turns than when using it without prime. However, learners, at least in some instances, equally reused incorrect structures. In sum, the study showed that structural alignment occurred also in non-scripted, authentic interaction between L2 learners and that the types of structures depended on the task content and goals. Structural alignment was furthermore linked to accurate production provided that the preceding structure was also accurate.

While more research is needed on the effect of alignment on FL and L2 development, the current body of research suggests that learners benefit from syntactically and phonologically aligning to a more proficient speaker producing the target-like form. For structural alignment, the effect might be pronounced for difficult, infrequent structures. With respect to peer interaction, it needs to be noted that learners might equally align to erroneous forms when interacting with other L2 learners. The study by Dao et al. (2018) finding the highest rate of alignment for self-alignment, shows that, outside the laboratory context, alignment also plays a role in terms of automatization.

2.1.3 Summary and Rationale for Choice of Theoretical Framework

This subsection first summarizes the concepts and empirical evidence of the cognitive-interactionist research presented above. It then argues for the choice of the cognitive-interactionist approach as theoretical framework for this study and points

out the limitations of using a cognitive rather than sociocultural approach to studying EFL dialog system-interaction.

To sum up the role of interaction for FL and L2 acquisition from a cognitive-interactionist perspective, theory and empirical evidence suggest that interaction promotes language development by providing learners with opportunities for input, output, and interactional features. These interactional components are understood to act and interact in the following ways in the acquisition process:

- The linguistic input that the interlocutor produces during interaction provides the learner with positive evidence to build hypotheses about the FL/L2, given that the input is comprehensible to the learner. In interaction, comprehension may be achieved by modifying the input through negotiation of meaning. Effective input processing might be enhanced when the noticing of linguistic features is enhanced, for instance through increased salience.
- The output produced by the learner during interaction supports FL and L2 learning in multiple ways: it enhances fluency through automatization; it prompts the learner to become aware of gaps in their FL/L2 and directs their attention to linguistic information in the input needed to fill them; it provides the learner with opportunities to test their hypotheses about the FL/L2 and ‘try out’ language including modifying it to make it more target-like; it allows learners to reflect on their language use through language.
- Interactional features that occur specifically in learner interaction promote the processing of input and output in the acquisition process, thus explaining the essential role interaction plays in FL and L2 learning: negotiation of meaning makes input comprehensible, draw the learner’s attention to specific linguistic features as well as form-meaning relationships, serve as a source of feedback, and prompt learners to modify their output. Through feedback, learners become aware of their errors and, in the case of input-providing feedback, may be provided with the correct form. Feedback may also prompt learners to modify and repair their output. The benefits of feedback appear to be most pronounced in explicit, output-prompting feedback types, at least in meaning-oriented classrooms with lower-proficiency learners. Lexical, structural, and phonological

alignment contribute to the automatization of linguistic features and more accurate and complex production through repeated exposure and production.

With its well-established theoretical claims and its substantial body of empirical research in different learning contexts, including computer-mediated and, to some extent, dialog system interaction, the cognitive-interactionist approach was considered a valid theoretical framework for the current study on learner-system interaction and its potential for EFL learning. Furthermore, when planning this study, common anecdotal experience by dialog system users characterized dialog system interaction as filled with misunderstandings and attempts to resolve them, for example through rephrasing or aligning to the system's language use. So, the focus of the cognitive-interactionist approach on concepts such as enhanced input, negotiation of meaning, or alignment along with its proven methods for identifying and analyzing them, appeared well suited to describe the nature of the above-mentioned interactional characteristics.

However, there are limitations to the cognitive-interactionist approach: Firstly, learner-system interaction, as in all interaction humans engage in, has social and contextual dimensions that are out of scope of the cognitive-interactionist approach. It can be expected based on research on human-human FL interaction conducted from sociolinguistic or sociocultural perspectives that factors such as, for example, the institutional setting or the learners' psychological states, beliefs, and cultural backgrounds, impact on the nature of the interaction and its learning outcomes. In this study, additional descriptive data from learner language tests, questionnaires, and interviews provide some initial insight into possible relevant factors that come into play in learner-system interaction; however, these will need to be researched further with an explicit focus on these social and psychological aspects.

Secondly, the approach does not include an examination of interactional features such as scaffolding, collaborative dialogue or language-related episodes (LREs) which are conceptualized and researched by the sociocultural approach to FL and L2 learning. While the cognitive-interactionist approach focuses on learning as an individual cognitive effort, the sociocultural approach views language learning as a collaborative process that occurs during interaction with a more skilled speaker (e.g., a

teacher or another learner) (Lantolf, 2011). In interaction, the more skilled speaker provides scaffolding for the learner who, in this way, manages to co-construct new knowledge and skills. Learning may also be mediated through LREs in which peers explicitly discuss their language use, e.g., to collaboratively arrive at a grammatical form that they were incapable of producing on their own (Lambert et al., 2019). At the current stage of technical development, no dialog system known to the author is capable of responding situationally and individually to a degree that is required for such collaborative dialog. However, the concept might become more pertinent in the future as the technology advances and the inclusion of a sociocultural approach to studying learner-system interaction could become more relevant.

As a related third limitation, the cognitive-interactionist approach is mainly concerned with the production and development of discrete linguistic features (i.e., lexical items or grammatical forms) rather than higher-level discourse features considered by discourse and conversation analytic and sociolinguistic FL and L2 research. Thus, cognitive-interactionist research excludes conversational elements such as, for example, speech acts, discourse functions, coherence and cohesion, or behaviors such as turn taking or discourse management. The different foci of the cognitive-interactionist approach, on the one hand, and a conversation analytic approach, on the other hand, becomes evident in their different ways of conceptualizing and analyzing the interactional feature *negotiation of meaning*. As shown above (see section Negotiation of Meaning), the cognitive-interactionist approach considers negotiation of meaning in terms of its function in providing learners with data and feedback to establish and revise their understanding of how linguistic forms relate to and express meaning. Instead, conversation analysis conceptualizes and analyzes negotiation of meaning in terms of its grounding function, in other words, how interlocutors use negotiation to establish and update common ground. Clearly, discourse analysis provides a useful additional framework to better understand FL learner-system interaction, particularly to tap into aspects such as, for example, the use of speech acts and discourse function, coherence and cohesion, which are highly relevant aspects of FL pragmatic competence and, accordingly, frequently studied by FL interactionist research. However, an evaluation from the discourse analytic approach is out of scope of the present analysis. To somewhat counterbalance

this limitation, some discourse elements related to learners' development of pragmatic competence were included in the analysis under research question 2: On the one hand, the accuracy assessment of the learners' task performance included an evaluation of the appropriateness of the speech act *request*. On the other hand, the interactions were evaluated in terms of their communicative adequacy which accounts for discourse structure and stylistic appropriateness. These concepts will be introduced in the following sections.

2.2 Task-based Interaction

In school-based FL learning and teaching, interaction is mediated by tasks. This includes EFL learning and teaching at commercial VET schools, where this study is situated. The reasons for the use of task-based interaction in FL classrooms will be outlined in the following paragraphs.

Numerous definitions exist of what constitutes a task in L2 and FL learning (see for example Ellis, 2003; Kim, 2015; Pica et al., 1993; Swales, 2009). Synthesizing 14 different definitions and descriptions from English, French, and German FL learning literature, Thonhauser (2016) has identified a list of features that characterize tasks in the current FL learning and teaching discourse. According to this synthesis, a task is an activity that:

- has a defined (communicative) goal such as, for example, the production of a text or the realization of a performance;
- has a focus on meaning, i.e., learners engage with meaningful content when carrying out the task;
- is relevant to the learners' experience, i.e., relates to the 'real world';
- has a focus on learners as agents, i.e., develops the learners' autonomy;
- enables authentic use of language by engaging learners in an activity that they experience as relevant and meaningful; and
- enables the development of language competence by enabling learners to focus on form, i.e., become aware of linguistic forms.

The following develops the argument for the need to focus on task-based interaction when investigating learner-system interaction in the context of commercial VET school EFL learning and teaching and introduces the additional theoretical concepts that informed this study, specifically task design and performance with regard to research question 2. First, it will be explained what makes tasks and task-based interaction suitable for school-based FL learning. This is followed by a description of how a task's efficacy is assessed via evaluating the learner's task performance through analyzing its interactional features, its complexity, accuracy, and fluency, and its communicative adequacy. It is considered relevant to introduce these concepts as, on the one hand, they informed the design of the task used in this study. On the other hand, the concepts were applied to analyze the resulting interactions in terms of research question 2 focusing on complexity and accuracy, as well as communicative adequacy of the learners' performance. Finally, the relationship between task performance and task efficacy mentioned above builds the foundation for discussing the task's potential for FL classroom learning and teaching based on the findings on learners' task performance.

2.2.1 The Role of Task-based Interaction in FL Classroom Learning and Teaching

Tasks are a useful tool to facilitate interaction in FL classrooms for two reasons. As a first reason, tasks may be used as "vehicles" (Kim, 2015) to elicit the kind of interaction that promotes FL learning, as claimed by the cognitive-interactionist approach. As explained above (see section Cognitive-Interactionist Approach), according to the cognitive-interactionist approach, FL learners and their interlocutors engage in interactional work such as negotiation of meaning, feedback, modified output, alignment, and so on to generate comprehensible input and output. These interactional features not only promote mutual understanding between interlocutors, but, as the cognitive-interactionist approach further suggests, also facilitate FL acquisition. Naturally, the interlocutors' engagement in this kind of interactional work requires that there is meaning to be negotiated and a need for mutual understanding. This condition, in turn, is created by tasks, as they create the need for learners to

negotiate towards mutual comprehension in order to achieve a communicative goal (Ellis, 2003).

The second reason for the use of tasks in FL classrooms relates to the central role of tasks in communicative language teaching (CLT), an influential teaching approach that focuses on using language as a means of communication, placing the development of communicative competence at the center of FL learning and teaching (Richards & Rodgers, 2014). Communicative competence entered L2 and FL research from the field of sociolinguistics where Hymes (1972) conceptualized communicative competence to include aspects of communication and culture. One of the earliest and to this day most influential adaptations of Hymes' communicative competence to the field of L2 and FL research was the interpretation of Canale and Swain (1980). The model and its components have since been revised by other L2 and FL scholars, as for example, by Canale (1983), Bachman (1990), and Celce-Murcia et al. (1995). While these various adaptations of Canale and Swain's model propose slightly different construct definitions, they generally define communicative competence as the knowledge and ability to use language efficiently and socioculturally appropriately. According to the CEFR, communicative competence consists of the three components linguistic, sociolinguistic, and pragmatic competencies (Council of Europe, 2001a). Linguistic competence is defined as "the knowledge of, and ability to use, the formal resources from which well-formed, meaningful messages may be assembled and formulated" (Council of Europe, 2001, p. 109). It involves all dimensions of the language system "independently of the sociolinguistic value of its variations and the pragmatic functions of its realisations" including, for instance, lexical, phonological, and syntactical knowledge and skills. Sociolinguistic competence refers to "the knowledge and skills required to deal with the social dimension of language use" (Council of Europe, 2001, p. 118) and includes aspects such as, for example, the use of markers of social relations (e.g., greetings, use of address forms), politeness conventions, or language varieties (e.g., registers, dialects). Pragmatic competence is concerned "with the functional use of linguistic resources" (Council of Europe, 2001, p. 13) and includes the ability to perform communicative functions and the competence to sequence sentences to produce coherent stretches of language and according to "interaction schemata" (e.g., genres) (Council of Europe, 2001, p. 123).

The concepts of CLT and communicative competence have had a major influence on FL teaching and learning, notably in the European context as these concepts build the foundation of the CEFR. The CEFR has informed many FL school curricula, including the Swiss curricula for primary and secondary school, the commercial VET school curriculum, and other VET FL curricula.

With its emphasis on language as a means for communication, CLT situates learning in meaningful activities that aim to engage learners in authentic communication (Richards & Rodgers, 2014). In this learning environment, tasks and task-based interaction naturally play a key role (Ellis, 2003). As stated in the list above, FL tasks are characterized by enabling learners to use authentic language while focusing on meaning to achieve a communicative goal, making the tasks an inherently communicative language practice. In fact, the establishing of tasks in the language curriculum “grew out of the communicative language teaching movement of the 1970s and 1980s” and were first introduced into FL and L2 classrooms in the 1980s when language teachers called for the need to teach communicative competence (Ellis et al., 2020, p. 3).

The interrelation between task-based interaction and FL learning from the perspectives of the cognitive-interactionist approach and CLT and communicative competence can be summarized as follows: From a cognitive-interactionist perspective, interaction tasks are used as tools to elicit the kind of meaning-focused interaction that promotes the production of interactional features and thus the acquisition process. Furthermore, task-based interaction is a key ingredient to the CLT classroom as it develops the learners’ communicative competence by placing them in communicative situations requiring them to use language effectively linguistically as well as sociolinguistically and pragmatically. Due to the prominent role of CLT and communicative competence in FL curricula, tasks and task-based interaction take a central position in many FL classrooms.

These considerations explain the relevance of task-based interaction in the context of this study: On the one hand, being framed by the cognitive-interactionist approach, the study used an EFL task as part of the data collection to elicit interactions with the features investigated from this perspective. On the other hand, the study is situated in the context of EFL teaching and learning at a Swiss commercial VET school

which implements curriculum based on the CEFR and its concepts of CLT and communicative competence highlighting the role of tasks and task-based interaction.

2.2.2 Evaluating Task Efficacy

In order to evaluate a task's efficacy with regard to its two-fold function of 1) facilitating interactional features and 2) promoting the development of communicative competence, research commonly examines the learners' performance when carrying out a given task (Ellis, 2018). In terms of the cognitive-interactionist perspective, task efficacy is evaluated by examining the occurrence of interactional features in the learners' performance. Interactional features such as negotiation of meaning sequences or feedback may be identified, categorized, and quantitatively and qualitatively analyzed (Ellis, 2018). Based on a quantification and qualitative description of the found interactional features, assumptions based on the cognitive-interactionist theories can be made about the task's affordance for FL learning. Tasks that elicit a high degree of negotiation of meaning, for example, are considered beneficial in that they promote cognitive processes involved in acquisition.

A task's role in promoting the development of communicative competence, i.e., linguistic, sociolinguistic, and pragmatic competence, is examined by means of the performance aspects complexity, accuracy, fluency, and, more recently, communicative adequacy. The four concepts will be briefly defined in the following paragraphs followed by an explanation of how these concepts relate to the development of communicative competence and task efficacy as a rationale for measuring these aspects when assessing task performance.

Complexity, accuracy, fluency

L2 and FL research has investigated complexity using different definitions and operationalizations (Bulté & Housen, 2012). Reviewing the literature on complexity, Bulté and Housen (2012) define complexity as “a property or quality of a phenomenon or entity in terms of (1) the number and the nature of the discrete components that the entity consists of, and (2) the number and the nature of the relationships between the constituent components” (p. 22). According to the authors, L2 and FL complexity

consists of the three components “propositional complexity”, “discourse-interactional complexity”, and “linguistic complexity” (Bulté & Housen, 2012, p. 24). Propositional complexity refers to the density of information or idea units in a text, with a higher ratio of information or idea units per unit corresponding to higher complexity. Discourse-interactional complexity relates to the complexity of dialogic discourse in terms of the number and type of turns that a learner initiates and interactional moves they engage in. Linguistic complexity can be understood as both “global or system complexity”, i.e., “the number, range, variety or diversity of different structures and items», as well as “structure complexity”, i.e., the complexity of the individual syntactic, morphological, phonological, or lexical features (Bulté & Housen, 2012, p. 25). Structure complexity of individual features concerns both functional complexity (i.e., the meaning-form relationship of linguistic features) and formal (i.e., the structure and components of linguistic forms).

Accuracy in L2 and FL research has been used to mean the degree to which the learner’s performance conforms to the norms of the target language (Housen et al., 2012; Housen & Kuiken, 2009; Pallotti, 2009). High accuracy corresponds to the target-like use of the language, i.e., the absence of deviations, usually characterized as “errors”, from L2/FL norms (Michel, 2017; Polio & Shea, 2014).

As has been pointed out by Housen et al. (2012) and Michel (2017), this definition raises the question of what constitutes norm in language use, as the normative standard may vary between, for example, different modalities (e.g., written, oral) or dialects. Therefore, when measuring accuracy, the norm needs to be defined and, for instance, agreed on between different raters.

Another question is whether the norm deviation is measured solely in terms of number (i.e., the number of errors is counted) or if the severity of the errors should be taken into account. This is based on the idea that some errors such as, e.g., typos or omission of punctuation, may be considered less severe than others such as, e.g., word disorder, omission of an article, or word disorder (Housen et al., 2012; Michel, 2017). For this reason, it has been proposed to calculate weighted accuracy scores, for instance by classifying errors in terms of severity or by assigning individual clauses accuracy scores (Michel, 2017).

Accuracy measures may include holistic rating assessing the overall impression of performance accuracy or analytic assessment of different aspects of the performance, which may involve either general measures such as, e.g., number of error-free clauses or errors per 100 words, or specific measures related to the focus of the investigation such as, e.g., target-like use of past tense forms (Michel, 2017).

Fluency is generally conceptualized as the ability or automaticity in verbalizing an intended meaning with native-like rapidity, smoothness, and hesitation (Housen et al., 2012). According to Tavakoli and Skehan (2005), fluency encompasses the three components (1) speed or rate of the language, measured by, e.g., number of words per minute; (2) silence or breakdown, e.g., amount, location, and duration of pauses; (3) repair, e.g., false starts, repetitions, and self-corrections.

While fluency has also been used in research on written production (e.g., assessed by measuring typing speed), it is most commonly used as a measure of spoken language (Michel, 2017). When measuring fluency in written production, it needs to be considered that the reiterative process of writing allows learners to plan, monitor, and edit their speech, which makes fluency more challenging to measure. At the same time, however, techniques to measure writing fluency such as the use of key-stroke logging software allows researchers to also review the process rather than the product alone of writing fluency (Michel, 2017).

Communicative Adequacy

Recent task-based research has begun to include communicative adequacy measures when assessing learners' task performance for the reasons that will be discussed below. Communicative adequacy of task performance has been assessed in different ways and using different terms and construct definitions in previous research. Thus, communicative adequacy has been considered in terms of *pragmatic appropriateness*, *communicative effectiveness*, *functional adequacy*, or *successful task completion*: For example, Youn (2015) assessed the pragmatic performance of learners in role-play interactions along five criteria (content delivery, language use, sensitivity to situation, engaging with interaction, and turn organization). Turner and Upshur (1995) examined the relation of grammatical accuracy and communicative

effectiveness (i.e., success in information transfer) rated on a task-related scale including criteria such as coherence and number of required elements. Sample and Michel (2014) added task completion (i.e., the number of differences found) to complexity, accuracy, and fluency measures when assessing learners' task performance in spot-the-difference interactive tasks. De Jong et al. (2012) measured performance in terms of functional adequacy (i.e., success in conveying information through speaking) on a scale with descriptors pertaining to the amount and detail of information conveyed, topic relevance, degree of formality, discourse type, and the ease with which the description could be followed.

As becomes evident from these definitions and operationalizations, communicative adequacy generally refers to the learner's ability to achieve the task's goals by using language that is adequate to the specific communicative context, thus incorporating both pragmatic as well as task completion aspects.

Rationale for Assessing CAF and Communicative Adequacy of Task Performance

As mentioned above, CAF measures are commonly used to investigate how a task impacts the learners' performance (Ellis, 2018). Underlying this is the theoretical assumption that the quality of the learners' performance in terms of CAF relates to the development of their FL or L2 competence (Skehan, 2003) as the performance's complexity, accuracy, and fluency is understood to reflect the underlying cognitive processes driving L2 and FL acquisition (Foster & Skehan, 1996). Thus, the use of complex language may indicate that learners are experimenting with language forms and thus expanding their linguistic FL/L2 competence. High accuracy levels may indicate that learners are focusing on form and applying their linguistic knowledge. A high level of fluency may indicate that learners are proceduralizing their knowledge. From this perspective, CAF in learners' task performance can be viewed as indicators of the cognitive processes promoting language development (Housen et al. 2012) and thus inform on a tasks' efficacy in facilitating the development of L2/LF competence.

In this context, it has been suggested that the three CAF components are interconnected and may interact in the process of FL/L2 production and development (Housen et al., 2012; Michel, 2017). Thus, Skehan's (1998) limited attentional capacity

model argues that FL and L2 learners are unable to allocate their unlimited attention to all three components during task performance due to limited information processing capacity. As a consequence, a given task may concentrate the learners' attention on one of the three performance areas: for instance, learners may attend more to fluency during a given task at the expense of accuracy and complexity. However, the rival cognition hypothesis (Robinson, 2001b, 2005) rejects the idea that attentional capacity is limited and instead suggests that learners can draw on multiple attention pools simultaneously. According to this view, increased task complexity pushes learners to produce more complex and accurate language as the learner experiences the need to express more complex ideas, both complexity and accuracy of their performance increase – possibly at the expense of fluency.

To sum up the role of measuring CAF in task-based interaction, it is understood that the levels of complexity, accuracy, and fluency of learners' task performance relate to L2 and FL development in that they reflect the cognitive processes involved in the L2/FL development. However, the relationship between CAF and L2/FL performance and development is likely interrelated in a non-linear way, as the three components interact in intricate ways. As a consequence, low or high measures of CAF components in task-based interaction may not automatically be viewed as beneficial or detrimental, respectively, for L2/FL development but may instead also be indicative of the components' mutual supportive or competitive influence on each other.

Moreover, more recent task-based research has argued that, when assessing the learners' performance in a task, a further focus should be on whether the learner successfully achieves the task goal and uses language that is adequate to the communicative context set by the task. Pallotti (2009) was among the first to argue that assessing lexico-grammatical aspects alone is insufficient when evaluating task performance, since learners may produce linguistically complex and accurate language while being functionally ineffective. In fact, as Pallotti (2009) further suggests, the learner's pragmatic choices could impact the CAF levels of their performance. For instance, a learner with limited proficiency may prioritize producing an understandable sentence that is communicatively effective over producing a sentence that has no errors.

So, on the one hand, communicative adequacy through its interrelation with CAF as the linguistic dimensions of performance may help to interpret CAF measures in task performance (Pallotti, 2009). On the other hand, communicative adequacy as a separate measure of task performance may indicate the degree to which a learner is successful in achieving the task's goal. The ability to communicate successfully in terms of the task's goal is central from the perspective of competence-based FL learning and teaching, as it may reveal how the learner engages with the task and its associated competences. As described above (see section *The Role of Task-based Interaction in FL Classroom Learning and Teaching*), in competence-based teaching, tasks are used with the goal of building FL competences (Keller & Reintjes, 2016), more specifically, communicative competence composed of linguistic, sociolinguistics, and pragmatic competences. Assessing the communicative adequacy of the learners' performance on a given task, i.e., the degree to which they respond adequately to the task's requirements and communicative situation, may serve as an indicator of the degree to which the task enables them to engage with the task and its targeted competences, specifically including pragmatic and sociolinguistic competences.

2.3 SCMC for Language Learning

As a subfield of computer-assisted language learning (CALL), computer-mediated communication (CMC) entered language classrooms in the 1980s (R. Kern & Warschauer, 2000; Warschauer, 1996a) and began to emerge in L2 and FL research from the 1990s (Payne & Ross, 2005). As Kern and Warschauer observed, the emergence of computer-mediated communication went hand-in-hand with a shift from a structural (1920s to 1950s) to a communicative perspective (based on cognitive and sociocognitive paradigms from the 1960s) of L2 and FL learning and teaching.

CMC subsumes two modes: synchronous computer-mediated communication (SCMC), i.e., technologically mediated real-time interaction between interlocutors as in, e.g., audio and video conferencing or text-based chat, and asynchronous computer-mediated communication (ACMC), i.e., interaction in deferred time as in, e.g., email or online forums (R. Blake, 2017). While it is true that communication technologies are “not inherently synchronous or asynchronous” (O'Rourke & Stickler, 2017, p. 3), as

email, for example, may be used synchronously, or chat may be used asynchronously, the technologies are generally used ‘as intended’ in L2 and FL research.

SCMC and ACMC may be text-, audio, or audio-visual-based as in videoconferencing (Ziegler & Phung, 2019). Although text-based SCMC involves “delayed” or “quasi-synchronous” communication as interlocutors produce the messages before, not while simultaneously, sending them, this subtype is included in the definition (O’Rourke & Stickler, 2017, p. 2).

SCMC in the L2 and FL classroom may involve communication between, for instance, learners and peers, between learners and teachers of the same class, or between learners and native speakers in an intercultural exchange, in one-on-one dyads as well as in groups with multiple participants (Chun, 2008; Michel & Cappellini, 2019).

Since the present study explores written SCMC between learners and peers, the following review will focus on research conducted on the written SCMC subtype. The review first presents the claimed benefits of SCMC for FL and L2 learning as variety of computer-mediated communication, followed by a survey of the relevant studies providing empirical evidence for these claims. The body of research on written SCMC is reviewed here for two reasons: first, as will be seen in the following paragraphs, it provides empirical evidence for the application of the cognitive-interactionist and task-based approach from FTF to computer-mediated interaction. This warrants the question whether the cognitive-interactionist and task-based claims and concepts may also be applied to the modally similar learner-system interaction. Along with the emerging evidence on learner-system interaction itself, which will be presented in the following chapter, the evidence on the effectiveness of SCMC thus supports the rationale for conducting the present study. Secondly, previous studies on SCMC informed the present study methodically, for instance, it guided the analysis of the SCMC data.

2.3.1 Claimed Benefits

From a cognitive-interactionist perspective, it has been hypothesized that written SCMC might offer the same benefits for L2 and FL learning as FTF interaction

and thus facilitate language development by providing learners with opportunities to receive input, produce output, and benefit from interactional features such as negotiation of meaning, feedback, and alignment.

This claim is based on the characterization of written SCMC as a “combination of writing and speech” that offers learners the opportunity of “text-based interaction” (Warschauer, 1997, p. 6). In other words, written SCMC shares the “real-time pressure” (O’Rourke & Stickler, 2017, p. 5) of FTF interaction, requiring learners to exchange information spontaneously and providing them with immediate feedback. At the same time, written SCMC appears in a textual format of letters and punctuation which are decoded visually by the interlocutor (Jonsson, 2015).

As unique characteristics of L2 and FL interaction through written SCMC compared to oral FTF interaction, research has identified, for instance, that turns commonly overlap due to the lag time between turns and the fact that messages are preserved in the chat (Smith, 2003). There are also longer pauses between turns due to, on the one hand, typing and message transmission and, on the other hand, the learners’ self-monitoring in which they edit their messages before sending them (S. Sauro & Smith, 2010). In their text production, learners have been shown to use shorter sentences, simplified syntax, symbols, and emojis and emoticons. They also tended to be accepting of surface errors resulting in frequent misspellings and used openings and closings optionally.

Due to its written nature, text-based SCMC has been suggested to provide learners greater opportunity for noticing and monitoring and longer processing and reviewing time for comprehension and production, compared to FTF oral interaction (Chun, 2008; Pellettieri, 2000; Smith, 2004; Warschauer, 1997), potentially leading to improved learners’ task performance (S. Sauro & Smith, 2010) and comprehension (Ziegler, 2016). Specifically, it has been suggested that written SCMC, due to its authentic contextualization, may lead to learners focusing on completing the task’s communicative goals (task completion, communicative adequacy) while encouraging them to produce linguistically complex and accurate language (Collentine, 2019). Furthermore, as SCMC exhibits written as well as oral features, it has been suggested that it might allow learning gains not only in written productive skills but also a transfer of skills to L2/FL speech (Chun, 2008). Others have instead highlighted the

affordance of SCMC to help learners develop communicative competence in the SCMC medium itself and advocate for an expansion of the construct of L2/FL linguistic competence (O'Rourke & Stickler, 2017).

It has also been suggested that, compared to FTF interaction, communication via a computer might provide learners with a less stressful environment, potentially leading to reduced language anxiety, i.e., learners' situation-specific feelings of stress, fear, or nervousness when using an L2 or FL (MacIntyre & Gardner, 1994) which may negatively impact learners' cognitive processing and willingness to communicate, thus inhibiting L2/FL learning (Cote & Gaffney, 2021). Especially less active students may be more encouraged to participate than in an FTF environment (Chun, 1994; Warschauer, 1996b).

2.3.2 Empirical Evidence

This subsection presents the research on written SCMC as it relates to the present study: The first part reviews studies on the quality of SCMC task performance in terms of CAF and communicative adequacy. The second part surveys research from the cognitive-interactionist perspective, focusing on the question in how far the benefits of interaction found in FTF interaction apply to written SCMC and what unique benefits written SCMC may provide to learners in terms of receiving comprehensible input, producing output, and engaging in interactional features. The final part presents evidence on the direct effect of written SCMC on language development.

Earlier SCMC studies explored the general quality of learners' performance during written SCMC. In one of the earliest studies on written SCMC, Chun (1994) examined the written transcripts of 14 SCMC learner discussions that took place in two L2 German classes over the course of two semesters. The author found promising features such as high learner initiative, increasing ratio of complex over simple sentences over the course of the two semesters, the provision of peer feedback, and the use of functional and interactive utterances (e.g., questions, statements and imperatives, requests, greeting and leave-taking). Similarly, Kern (1995) explored the quantity and quality of learner production in an SCMC group discussion of L2 French

university students in comparison to an oral FTF discussion. The results revealed a larger amount of language produced in written SCMC compared to the FTF discussions, measured as number of turns, sentences (T-units in the FTF discussions), and words. The author accounts for the lower amount of language production in the oral discussion in part due to the instructor's higher participation and the increased difficulty of the task compared to the SCMC discussions. SCMC output also showed a higher variety of grammatical structures (e.g., different verb tenses and moods, subordinate constructions, *if* and negative clauses) and discourse functions (e.g., greetings, assertions, questions). So, the initial findings of these earlier descriptive studies on SCMC were that the mode was characterized by greater student participation leading to increased output quantity and a variety of learning-facilitating features in the discourse.

More recent studies focusing on learners' task performance produced more nuanced findings on the quantity and quality of the learner production. For example, Fitze (2006) reported no statistically significant difference in the total number of words of written SCMC and FTF interaction. However, SCMC led to lexically significantly more complex language, in terms of greater lexical range. In contrast, Sauro (2012) found no significant difference in the lexical or syntactic complexity of learner FTF spoken interaction and written SCMC. However, further analysis revealed that some learners generated more complex language in one of the two modes, indicating that learner factors such as length of residency in the L2 culture, time on task, and familiarity with the English keyboard might have moderated the results.

More recently, Kim (2017b) compared the production of articles and question forms in FTF and written SCMC. The participants interacted in dyads on three communicative tasks using both written SCMC and FTF conversation. In the transcripts, articles were coded for correct or incorrect use, questions were coded according to their developmental stage. The findings revealed that written SCMC discussions produced more high-level, i.e., more complex question forms and a higher accuracy rate in the use of articles than FTF interactions. The author suggested that SCMC may force learners to produce complete and more advanced question forms to communicate clearly and efficiently, as they cannot rely on paralinguistic cues and need to manage split and overlapping turns. Furthermore, higher accuracy might have

been a result of an increased level of attention to forms in the written SCMC, compared to oral FTF interaction.

To sum up the studies on aspects of learners' task performance in terms of overall output quantity as well as complexity and accuracy, the findings point to a general advantage of written SCMC, compared to FTF oral interaction, however, they appear to be moderated by individual learner factors. Higher levels of complexity and accuracy in written SCMC might be a result of the mode's reliance on communicative clarity and efficiency and increased attention to form.

A few studies instead focused on the pragmatic appropriateness as a specific aspect of accuracy in SCMC task performance. These studies typically investigated the occurrence and nature of specific pragmalinguistic forms such as address forms or aspects at the discourse-level such as discourse functions (e.g., openings and closings) and speech acts (e.g., requests, refusals). Key studies on each of these aspects of pragmatic appropriateness will be presented below grouped by aspect.

With respect to address forms, Belz and Kinginger (2002) examined one German and French learner each who engaged in email interaction and written SCMC with a native speaker of the respective target language. The study used microgenetic analysis to observe the learners' developments in French or German, respectively, informal (T) and formal (V) pronouns of address. The data showed that the German learner moved from initially using T and V forms in free variation to using the appropriate T form consistently towards the end of his partnership and after his interlocutor requested him to address her by the T form. Similarly, the French learner gradually shifted from using T and V forms freely to using the T form more consistently after receiving explicit peer correction. So, this study demonstrated, on the one hand, that written SCMC provides learners with opportunities to receive feedback from their conversational partners and, on the other hand, that learners may develop the pragmatic appropriateness of their production through written SCMC interaction.

Tudini (2007) also explored address forms by investigating negotiation of meaning triggered by intercultural (i.e., content and pragmatic) issues in Italian learner-native speaker written SCMC interactions. Of the 37 intercultural negotiation episodes found in the data, 21 were intercultural-pragmatic negotiation which involved

native speaker corrective feedback on the learners' use of formal *lei* rather than informal *tu* which was considered appropriate by the native speaker. The study thus showed that written SCMC can provide learners with opportunities to explore and receive feedback on the use of terms of address in the target language. It furthermore exemplifies that the appropriateness of form, here form of address, is negotiated between the interlocutors in SCMC.

A similar observation was made by Ware (2005) who examined the interaction between learners from Germany and the United States through SCMC in German and English. The author found that the nature of the writing, including its formality, must be negotiated by the interlocutors as their individual expectations and norms for the task and for online communication in general differed.

Kim and Brown (2014) investigated the use of Korean address forms in four L2 learners' synchronous and asynchronous CMC interactions. The authors observed that the limited pragmatic competence of the low-proficiency learners restricted the appropriateness of address forms; however, they did use appropriate address forms such as kinship terms or titles when interacting with their interlocutors, especially the intermediate-level participants. Furthermore, the CMC environment enabled the learners to develop their competence in using the target forms by providing them with opportunities to use them in interaction and receive feedback on their use.

In sum, the findings produced by studies on the appropriate use of address forms in written SCMC so far indicate that learners appear to differ in the appropriate use depending on their proficiency but also individual expectations of formality for chat communication. The findings further indicate that SCMC can assist learners in developing their understanding of the rules of addressivity and a more pragmatically appropriate production, particularly due to feedback provided by the interlocutors.

With regard to the learners' production of discourse functions in SCMC, the early study cited above by Chun (1994) found that L2 German learners produced a number of discourse functions during written SCMC such as questions, statements and imperatives, requests, greetings, and leave-takings. Similarly, the above-mentioned study by Kern (1995) investigated the occurrence of discourse functions in L2 French written SCMC compared to FTF oral discussions. The author found a greater variety and number of discourse functions in written SCMC compared to the oral discussions.

For instance, SCMC produced four times as many assertions as FTF discussion and seven times more student questions which were, in addition, more varied and more complete compared to the FTF context.

More recently, Abrams (2008) focused on opening and closing sequences as well as topic initiations and developments produced by German learners engaging in class discussions through written SCMC. In the chat scripts, it could be observed that openings were always initiated by the students rather than instructors and involved extensive greeting sequences with each participant providing a greeting, mostly in an informal, but often in a formal style. The participants also performed extensive closing sequences with the first student asking for permission to leave the discussion in elaborate requests, followed by all the other students saying goodbye and stating the reason for leaving. Topic initiation and development were also controlled by the students rather than the instructor. The author observed that students' responses were more elaborate than they typically are in classroom discussions and the communication seemed more genuine, possibly because of a reduced cognitive demand in the written context as well as the more 'democratic' environment. Furthermore, the topic development was highly dynamic, with topics being commented on by several students or multiple themes being discussed simultaneously. Overall, SCMC in this study triggered authentic learner-learner interaction and improved pragmatic performance: the conversation and topics were learner-initiated, learners focused on the topic rather than the activity and performed a variety of elaborate discourse functions which are typically not accomplished in FTF classroom interaction.

With regard to speech acts, *requesting* and *refusing* have been of particular interest due to their face-threatening potential if executed inadequately. For example, Sykes (2005) compared the production of refusals in L2 Spanish learners interacting through oral SCMC, written SCMC, or FTF communication. The learners participated in group discussions and pre- and post-test role play tasks eliciting refusals. The produced refusals were subsequently analyzed in terms of number, complexity and variety of strategies that were used to mitigate the speech act. The results showed that the FTF group used the highest amount of support moves, while the written SCMC group used more than the oral SCMC group, although the latter difference was not

statistically significant. Furthermore, all three groups showed an increased variety of strategies used in the post-test, with the written SCMC group outperforming the other two groups in terms of the variety and complexity of the refusal strategies. The author suggested that the differences might have been due to the specific nature of interaction environment. Thus, FTF communication put the highest “pragmatic pressure” (2005, p. 418) on learners while the written SCMC context required more explicit communication due to the lack of intonation or body language that could be used to intensify the mitigators.

Research on the communicative adequacy of SCMC learners’ task performance and claims that the authentic contextualization of SCMC tasks may promote task completion is still scarce. As one exception, Collentine (2019) investigated the relationship between task completion, an aspect of communicative adequacy, and linguistic complexity and accuracy in learners’ written SCMC performance. In the study, Spanish learners solved two cases by interacting through pre-written dialogues with characters in a 3D game. The participants subsequently interacted in dyads through written SCMC with the goal of agreeing on the correct solution of the cases. The chat scripts were coded in terms of task completion (i.e., learners reached consensus about the solution of the case) as well as three linguistic complexity measures (i.e., TTR, lexical density, clauses per c-unit) and one accuracy measure (i.e., percentage of error-free clauses). Regression analysis suggested that higher task completion was associated with increased syntactic complexity but decreased accuracy. Conversely, learners with low task completion scores produced language with low complexity but few errors. However, this pattern could only be observed in learners with high and low task completion scores but not those with average scores. The author interpreted that learners’ focusing on communicating to solve the task encouraged them to produce complex language which, however, had low accuracy. Furthermore, the author suggested that virtual environments are a suitable context for learners to focus on meaning rather than form while communicating.

To summarize the evidence on pragmatic appropriateness in written SCMC, the learners’ pragmatic development may benefit from interacting in this mode, as it gives them the opportunity to produce a great variety of target forms and functions and to receive feedback from their interlocutor. As a challenge associated with SCMC, the

pragmatic performance by learners appears to be somewhat negatively impacted by the chat context, possibly because learners tend to associate the mode with more informal, private communication and feel less pressured to produce pragmatically appropriate language compared to the FTF context. Similarly, the findings on communicative adequacy in written SCMC suggests that the environment might lead learners to focus on meaning more than form, compared to FTF interaction. However, a more pragmatically appropriate use of language in written SCMC may be negotiated through interlocutor feedback.

To investigate the application of the cognitive-interactionist approach to the written SCMC context, some studies have focused on the learners' output monitoring and subsequent production of modified output in written SCMC. For example, Sauro and Smith (2010) explored the impact of output production during written SCMC on linguistic complexity and lexical diversity. The participants, university German learners, interacted in dyads on a communicative task through written SCMC. With the help of screen recording videos, the chat transcripts were coded for complexity measures (i.e., c-units, clauses, instances of the use of grammatical gender) as well as the evidence of post-production monitoring (i.e., text insertions and deletions before sending the message). The results revealed an overall higher linguistic complexity measured by clauses per c-unit and number of grammatical gender uses and lexical diversity, measured by type-per-token ratio in text that was typed after deletions or corrections, which can be considered as a result of learner monitoring. Thus, this study indicates that SCMC may not only lead to greater linguistic complexity because of overall increased production, as suggested in earlier studies, but also because learners benefit from opportunities for monitoring. This contrasts with the suggestion above that the SCMC mode might inhibit focus on form, indicating that the learners' focus on linguistic form can be enhanced by the specific setting (e.g., task, instruction) in which the interaction takes place.

Other studies investigated feedback in written SCMC as well as the claim that it might be better noticed by learners than in FTF interaction. For example, Sotillo (2005) explored feedback in learner-peer as well as learner-native speaker SCMC. Three native speakers and three ESL learners interacted on five communicative tasks using the oral and written modalities of a chat application. Chat logs and transcripts of the oral data

were combined for analysis and coded for corrective feedback type and characteristics. The results revealed that feedback occurred in 65 of 159 opportunities. Furthermore, learner-peer interactions produced more feedback opportunities and more actual feedback episodes than learner-native speaker interactions. The authors suggested that native speakers might have been more focused on meaning than form and wanted to avoid correcting their interlocutors out of politeness and to prevent discouraging them from using their second language. The most frequent type of feedback was indirect feedback (i.e., recasts, confirmation checks, clarification requests, and comprehension checks) and included corrections of vocabulary, grammar, spelling, and (in a single instance) pronunciation errors.

Similarly, Sauro (2013) explored the potential of feedback in written SCMC between English learners in Sweden and teacher trainees in the United States. The US teacher trainees were given the task to focus on specific target items and provide feedback to the learners during interaction. Qualitative analysis of the chat scripts showed that feedback was often enhanced by the use of special characters increasing the salience of the feedback and error. Furthermore, the teacher trainees elaborated their feedback by adding conversational continuants in order to mitigate its directness and provide additional information on the error such as a possible reformulation. Overall, the author observed that the consistency of feedback varied depending on the specific target form and the teacher trainee's background. There was also learner uptake, commonly in the form of acknowledgements of the feedback (e.g., "Okay, thanks") or, rarely, by using the target form in isolation or in context in a subsequent utterance. However, uptake tended to be limited by the specific feedback type and the writing context. Thus, the uptake of target forms provided in recasts seemed to be considered redundant by learners since the target form remained visible on the screen.

So, while task-based written SCMC does provide opportunities for learners to receive different forms of feedback in response to different error types, the mode might dissuade interlocutors to give direct feedback and prevent learners from uptake and producing the target-like form in their own output following the feedback.

Studies focusing on the direct effect of feedback on the development of forms has provided further insight into the specific characteristics and efficacy of feedback in written SCMC. Loewen and Erlam (2006) reported no effect for feedback (i.e., recasts

or metalinguistic prompts) provided in written SCMC on the development of English past tense in elementary L2 English learners. According to the authors, possible reasons might have been that the learners were developmentally not ready for the acquisition of the target form, that the feedback was not provided immediately due to message overlap, and the lack of uptake.

Sauro (2009) investigated the impact of recasts and metalinguistic prompts in written SCMC on immediate as well as long-term gains in L2 grammatical knowledge. The participants interacted with native speakers through written SCMC under three different conditions: treatment group one received feedback in the form of recasts, treatment group two was provided metalinguistic prompts, and the control group did not receive any feedback. Language gains with regard to the target form were measured through a pre-post-test design with an immediate as well as delayed post-test. The differences between pretest and immediate post-test scores showed that the feedback provided was effective in improving the participants' L2 knowledge on familiar items in the short-term, compared to the control group that showed no improvement; however, only the effect of metalinguistic prompts was significant.

Yilmaz and Yuksel (2011) provided additional evidence for the benefits of feedback in written SCMC, focusing on recasts. In their study, the participants, native English speakers with no Turkish knowledge, were instructed on two Turkish grammatical structures which were subsequently elicited in two treatment tasks, one carried out through written SCMC and the other through FTF interaction with the researcher who provided recasts for the target structure. Acquisition of the target forms was measured through an oral post-test. The results revealed that learners scored significantly higher on the post-test when they received SCMC recasts compared to FTF recasts. The authors propose as one possible explanation that noticing might be increased during written SCMC compared to oral FTF interaction.

Lai and Zhao's (2006) study provided partial support for this claim. The study compared learners' noticing of feedback in learner-peer FTF and written SCMC interaction. The participants, ESL learners, interacted in dyads on two communicative tasks which were designed to elicit feedback. On the day after the interaction, the participants recalled their noticing of the feedback in stimulated recall. Analysis of the data revealed that, overall, learners noticed the feedback better during written SCMC

than FTF interaction. Breaking the noticing down by feedback type shows that noticing of recasts was similarly low in both modes (18% FTF; 10% SCMC), while negotiation of meaning was better noticed in the written SCMC mode (18% FTF; 31% SCMC) although the difference was not statistically significant.

In contrast, Smith and Renaud (2013), using eye tracking technology, found that learners fixated on recasts in 72% of the instances, indicating a high degree of noticing of feedback in written SCMC. The study furthermore found a significant relationship between noticing and success in producing target lexical and grammatical forms in a post-test.

Gurzynski-Weiss and Baralt (2014) also investigated learner noticing of feedback in written SCMC and FTF mode. Spanish learners interacted with the researcher on two dyadic tasks in each of the two modes and subsequently recalled their noticing of the feedback during a stimulated recall session immediately after the task interaction. They found that learners reported in stimulated recall equal degrees of noticing of feedback (e.g., recasts, negotiation of meaning, metalinguistic feedback) in written SCMC and FTF interaction (68% FTF; 71% SCMC).

Li's (2010) meta-analysis on the effectiveness of feedback based on six studies confirmed that feedback given in written SCMC does have a positive effect on language development. The study found an overall medium effect of $d = 0.62$ for feedback in the SCMC mode. However, no significant differences between feedback in SCMC and feedback in oral FTF interaction was found. Thus, while the study suggests no advantage of SCMC feedback over FTF feedback, it highlights the efficacy of feedback in both communication modes.

In sum, research to date has produced evidence that written SCMC does provide learners with opportunities to receive feedback. Some of the specific characteristics of feedback in the SCMC mode such as enhanced noticing might increase its effectiveness, while others, such as lack of direct feedback, reduced uptake and modified output, and lack of immediate feedback might diminish the positive effect. However, evidence for these claims is still very limited and inconclusive. Overall, an initial meta-analysis of six studies has shown a medium effect for feedback in SCMC but no advantage over FTF interaction.

SCMC research has also investigated the interactional feature negotiation of meaning in computer-mediated interaction. Early studies explored whether negotiation of meaning is present in SCMC and described the episodes' characteristics in comparison to those occurring in FTF interactions. For example, Blake (2000) explored the occurrence of negotiation of meaning in written SCMC interactions of L2 Spanish learner dyads. The results revealed that learners negotiated meaning, predominantly on lexical non-understanding and, rarely, on phonological, morphological, and syntactic issues.

Lee (2001) explored the specific characteristics of negotiation in SCMC further. In this study, intermediate L2 Spanish learners interacted in dyads on a number of discussion topics through written SCMC once a week over the course of a semester. Analysis of the chat transcripts revealed a number of interactive strategies that the learners used to facilitate comprehension of input and output including, for instance, comprehension checks, confirmation checks, clarification requests, use of English, and self-corrections. The author observed that the responses that the learners produced in reaction to an indication of non-understanding tended to be brief and involved repetitions or the change of individual segments or words rather than elaborate modifications, possibly due to the learners' low proficiency. Furthermore, the learners' negotiation was limited to negotiation of meaning rather than negotiation of form, with many errors being ignored by the interlocutors in order to drive the conversation forward without interruptions. On the other hand, a considerable amount of monitoring and self-correction with respect to both lexical items and grammatical structures occurred. The author concludes that learners used negotiation of meaning and other strategies in similar ways as during FTF interaction, and thus facilitated mutual understanding.

Smith (2003) similarly investigated the nature and occurrence of negotiation of meaning in written SCMC. Learners interacted in learner-peer dyads on two jigsaw tasks and two decision-making tasks. Negotiation of meaning episodes were identified in the chat scripts based on Varonis and Gass' (1985) model, and each component (i.e., trigger, indicator, response, reaction to response) was coded according to its specific type based on categorizations used in previous research. The results revealed that learners negotiated about one third of the total discourse and negotiated significantly

more in the decision-making compared to the jigsaw task, possibly because the target lexical items were more salient in the former task type. Analyzing the specific nature of the negotiation routines, the author found that most sequences were concluded with a reaction to response move such as “ok” or “good”, indicating that learners might have felt more compelled to explicitly complete the routine in SCMC as many of the para- and nonlinguistic aspects of FTF interaction are lacking in SCMC. Similarly, most indicators explicitly identified the trigger, possibly because a general indication of non-understanding could be ambiguous to the interlocutor. Responses were almost always realized as rephrasals/elaborations, while responses with minimal modifications and trigger repetitions were rare. The author furthermore proposed an extended model based on the SCMC data. This extended model includes “split negotiation routines” where due to a lack of strict turn adjacency, the indicator or response move occurred at a later stage in the discourse. It also includes two additional phases following the reaction to response move, “confirmation” and “reconfirmation”, which serve to conclude the negotiation routine in cases where initiator and respondent of the negotiation routine confirm and reconfirm, respectively, understanding. The study thus showed that negotiation of meaning does occur in written SCMC and displays specific characteristics and a particular routine pattern which is represented more accurately by an extended model.

Later studies investigated which communicative mode (FTF, written SCMC, spoken SCMC) elicited the highest amount of negotiation. For instance, Fernández-García and Martínez Arbeláiz (2003) compared the negotiation of meaning of learner-peer, learner-native speaker, and native speaker-native speaker dyads in oral FTF interaction and written SCMC. The participants were native speakers and learners of Spanish at a US university. All dyads completed two communicative tasks, one through written SCMC and the other through oral FTF interaction. The results revealed that the learner-native speaker group negotiated significantly more than the other two groups; however, only during FTF interaction. The authors suggest that learners might have had more difficulty understanding the native speakers’ input than that of other learners’ for linguistic as well as sociocultural reasons. The learner-native speaker group also negotiated significantly more during FTF interaction than written SCMC, while there was no significant effect for mode in the other two groups. According to

the authors, the written mode might have facilitated the understanding and allowed learners for more processing time, leading to fewer communication breakdowns and thus negotiation episodes. Thus, the study provides some evidence that learners might negotiate more with native speakers than with other learners during FTF interaction, and that with respect to this type of dyad there might be increased negotiation in FTF interaction compared to written SCMC.

Similarly, Loewen and Wolff (2016) found a lower number of negotiation of meaning episodes in written SCMC compared to oral SCMC and oral FTF interaction. Dyads of university ESL learners were assigned to one of the three communicative modalities and completed three different interactive tasks. The chat scripts and transcriptions, respectively, were coded for negotiation of meaning (i.e., confirmation checks, clarification requests, comprehension checks) as well as recasts and language-related episodes (i.e., segments in which learners reflect on their language use when performing a collaboration task). The results indicated that there were significantly fewer confirmation checks and language-related episodes in written SCMC compared to the other two modalities. The authors argue that written SCMC might not offer the same opportunities for negotiation of meaning as, on the one hand, the information remains on the screen, and, on the other hand, some learners tended to present all of the information at once without leaving time for the interlocutor to interact between turns.

While written SCMC appears to produce fewer negotiation episodes compared to oral interaction (FTF and oral SCMC), there is some evidence from at least one study that point to the potential advantage of written SCMC in facilitating noticing during negotiation sequences. Thus, Yuksel and Inan (2014) explored the effect of communication mode (FTF interaction vs. written SCMC) on the amount of negotiation of meaning and the degree to which they are noticed by the learners. EFL learners at a Turkish university completed two different jigsaw tasks, one in FTF interaction and the other through written SCMC in learner-peer dyads. Noticing was assessed through the learners' retrospective comments on negotiation sequences during stimulated recall interviews. The results indicated that although there were fewer negotiation of meaning episodes in written SCMC than oral FTF interaction, learners noticed these sequences significantly more frequently.

Finally, Smith (2004) explored the direct effect of negotiation of meaning on lexical acquisition in written SCMC. In the study, ESL learners at a US university interacted with each other through written SCMC on communicative tasks once a week over a five-week period. Each interaction was followed by immediate and delayed productive and receptive post-tests to assess short- and long-term language gains. The results showed that negotiated lexical items were retained at a very high rate in the immediate and delayed receptive post-test (95% and 94%, respectively) as well as the immediate and delayed productive post-tests (85% and 81%, respectively). Furthermore, the post-test scores were significantly higher for lexical items that were negotiated compared to items that were described by a learner in the input instead (e.g., “I have a bouquet that is a lot of flower in the basket”); however, the higher score in the delayed productive post-test was not significant. Thus, this study provided evidence for the direct relationship between negotiation of meaning and L2 lexical acquisition in written SCMC.

To sum up the research on negotiation of meaning in written SCMC, there is evidence that negotiation of meaning does occur in this mode, albeit generally less frequently than in FTF interaction due to increased monitoring, longer processing time, and quicker turn-taking in SCMC. Negotiation of meaning occurrence has further been shown to be mediated by the interlocutors’ proficiency as well as task characteristics. SCMC research has also shown that negotiation of meaning occurring in this mode displays specific characteristics such as higher explicitness due to the lack of para- and non-linguistic aspects on the discourse. With regard to the benefits of negotiation of meaning in SCMC, there is some evidence that it involves increased noticing and has an effect on lexical acquisition.

More research has also involved studies on alignment in L2 and FL SCMC, some of which investigated the amount and specific nature of the feature as it occurred in the written SCMC mode, while others were interested in its effect on L2 and FL learning. In an early study, Uzum (2010) explored alignment in mixed-level ESL learner interaction through written SCMC. Qualitative analysis of the chat transcripts along with stimulated recall reports revealed that the participants aligned to each other on their choice of lexical items and grammatical structures. The author suggests that the

learners might have chosen to use lexical items and grammatical structures from their interlocutor's production in an effort to enhance understandability.

More recently, Michel and Smith (2017) investigated lexical alignment in L2 learner-peer written SCMC. During the interactions, the participants' eye movements and screens were recorded using eye-tracking software in order to measure whether there was overt attention to the interlocutor's production. The chat scripts were coded for exact lexical overlap of 3- to 10-word-units. These potential sources of alignment were then identified in the eye-tracking data and analyzed for gaze behavior. The results revealed that about two thirds of the aligned units were not accompanied by increased visual attention as measured by fixation data. The authors concluded that alignment in L2 SCMC might, to a large extent, be prompted by implicit automatic priming rather than explicit strategic behavior.

Michel and O'Rourke (2019) used the same procedure in a study on twelve German learners interacting through written SCMC with each other on a first and second task and a native speaking tutor on a second task. In addition, the researchers asked the participants to comment on the alignment during cued interviews. Results revealed overall higher alignment as well as higher eye fixation of aligned n-grams in the tutor chat than the peer chat. According to the authors, along with the learners' interview data, this finding suggests that when interacting with a native speaking tutor, learners chose alignment as a conscious strategy to use their interlocutor's input as a model for their own production. Furthermore, the degree of alignment generally decreased for higher-proficiency learners, possibly because more proficient learners are less reliant on their interlocutor's input to produce language. In sum, the study showed that both implicit, automatic as well as conscious, strategic processes played a role in the learners' alignment to their interlocutors' production, and that it was mediated by their own proficiency and the kind of interlocutor.

Michel and Cappellini (2019) included lexical as well as grammatical alignment in their study and further investigated how alignment was affected by different settings, i.e., videoconference or text chat and different target languages (Chinese, French, German). Data was collected in four French-Chinese dyads interacting through video chat in French and Chinese and in four German-speaking learner-peer and learner-native speaker pairs communicating via written chat. The chat conversations

and oral transcriptions, respectively, were coded for lexical and structural alignment. Lexical alignment was coded by initially identifying n-grams of three to ten exactly overlapping words that occurred at least twice in a given conversation. Structural alignment was coded by first identifying frequent structures, bottom-up as well as top-down following the structures identified by Dao (2018) (see above). Each first instance of a recurring item or structure would be coded as prime, a subsequent occurrence appearing within at most six turns (video data) or eight turns (written data) was coded as target, distinguishing “alignment to self” and “alignment to interlocutor”. Alignment and nonalignment scores were calculated by dividing the number of aligned and nonaligned, respectively, structures by the total number of occurrences in a given conversation. Comparing alignment in the two different CMC environments, the results indicated a larger amount of lexical alignment and higher lexical alignment scores in video conference compared to written SCMC, but more structural alignment and higher structural alignment scores in written SCMC than in videoconference. According to the authors, higher structural alignment in written SCMC might be due to increased salience in the text chat or to the fact that the written SCMC argumentative task might have yielded more alignment than the free conversations in videoconference. Another finding was that the type of interlocutor (L2 peer or L1 tutor) seemed to affect alignment to a smaller degree than the CMC environment did. Furthermore, structural alignment, which was mostly self-alignment, was generally most prevalent for frequent structures such as, e.g., *have* + complement, possibly because learners were more familiar with them and tended to avoid structures they did not know.

Kim et al. (2019) similarly compared structural alignment in two different modalities, FTF interaction and written SCMC. The study investigated alignment of the English stranded preposition construction in 114 Korean EFL learner interactions. During the interaction, the researcher supplied an object description either using the target construction or using a “filler”, i.e., an utterance not containing the target construction. After the learner guessed the described object, they described an item using a given verb. In the resulting interaction data, each stranded preposition construction in the learner production was coded as alignment. The results revealed higher production of the target structure following primes than when following fillers

in both the SCMC and FTF group, indicating that structural alignment does occur in both modalities. It was furthermore found that SCMC facilitated more alignment than FTF interaction. The authors suggested that this was due to the particular nature of the written mode which “tap[s] into different cognitive processes and strategies than those used during FTF interaction” (2019, p. 961) such as, e.g., during the possibility of scrolling up and down. The study also provided evidence for the learning-facilitating nature of structural alignment. Both experimental groups significantly outperformed the control group on the immediate post-test, and the SCMC group but not the FTF group also outperformed the control group on the delayed post-test. Moreover, the learners’ prior knowledge of the target structure but not their working memory capacity was found to be significant predictors of the amount of alignment occurring.

In a replication study of Kim et al. (2019), Kim et al. (2020) partially confirmed these findings. Testing the effect of alignment and individual differences (working memory, previous knowledge of target structure) on the acquisition of two question forms (*wh*-questions and indirect questions), the results confirmed that production of both question forms was more likely in the primed compared to the unprimed condition. There was also a modality effect, with SCMC producing significantly more alignment compared to the FTF context, but only for direct questions. According to the authors, indirect questions might have been less challenging for the participants to form and, therefore, the relative benefit of SCMC in providing more processing time compared to the FTF condition might have been less pronounced. Furthermore, learners might have tended to avoid indirect questions in SCMC to reduce typing. This finding suggests that the modality effects may be moderated by the specific target feature. Similarly, the results revealed a learning effect for direct questions, but not for indirect questions, possibly because learners preferred to use the shorter direct question form rather than the alternative long indirect question form requiring more typing during the production tests, thus masking any potential evidence. Finally, the study found no effect for working memory and prior knowledge, the finding on prior knowledge being contrary to Kim et al. (2019). According to the authors, this discrepancy could be explained with the fact that Kim et al. (2020) measured the prior knowledge variable only in terms of implicit rather than both implicit and explicit knowledge. In sum, this study provided further evidence for the effects of structural

alignment on grammar learning in both SCMC and FTF interaction. It also partially confirmed the relative advantage of SCMC over FTF interaction with regard to the amount of alignment, highlighting that the specific target feature might interact with modality.

Summarizing the research on the nature of alignment in written SCMC, there is evidence for the occurrence of lexical and grammatical alignment also in this interaction mode, as well as some initial evidence for the learning-facilitating effect of structural alignment. Moreover, structural alignment appears to be more frequent in written SCMC than in oral (FTF and SCMC) interaction. However, this advantage might only exist for forms that are relatively difficult yet known to the learners and effortless to type. As a further finding, alignment in written SCMC appears to involve both explicit/conscious and implicit/automatic learner strategies.

Finally, two meta-analyses have produced evidence for the positive effect of SCMC on L2 development. Lin et al. (2013) synthesized ten primary studies to investigate the effect of written SCMC on overall L2 acquisition compared to asynchronous CMC or FTF instruction. Lin et al. found a significant but small effect ($d = 0.33$) on overall L2 development for written SCMC compared to the other forms of interaction. An analysis of five contextual variables (i.e., L2 proficiency levels, treatment duration, target language, group division) suggested that the complex nature of SCMC task implementation might have impacted on the efficiency of the treatment and contributed to the relatively small comparative effect of SCMC.

Similarly, Ziegler (2016) meta-analyzed the effectiveness of SCMC on language development compared to FTF contexts. The study reported on 14 primary studies comparing the impact of written and oral SCMC modes (i.e., text chat, video conferencing, audio conferencing, or bimodal chat) and FTF interaction on L2 learning outcomes measured on immediate post-test scores. The results showed significant L2 improvements with large effects for both SCMC and FTF modes and a small comparative effect of SCMC ($d = 0.13$), indicating a slight advantage of SCMC over FTF contexts on overall L2 learning outcomes. The study also found a small advantage for SCMC over FTF interaction for the development of productive and written measures, whereas FTF interaction had a small advantage on receptive and oral measures. According to the author, the relative advantage of SCMC in promoting L2 production

development might be explained with the additional processing and planning time and increased attention in this mode, whereas FTF interaction might facilitate recognition of target forms in the input by offering learners additional input sources such as visual and verbal cues, thus promoting receptive skills. However, the author pointed out that the findings on the receptive and written skills must be interpreted with caution and be confirmed in further research due to the small sample size. The advantageous effect of SCMC for the development of written skills and of FTF interaction on oral skills suggests that the development of a given skill is promoted if the mode of the treatment is the same.

However, additional individual studies indicated that there might be transfer of skills in SCMC from one modality to another, specifically from written to oral modalities in written SCMC. For instance, Payne and Whitney (2002) compared the effect of written SCMC and FTF interaction on L2 oral proficiency in Spanish language classrooms. The participants were assigned to the experimental condition engaging them in FTF as well as written SCMC communicative activities or the control condition in which learners carried out the same activities in FTF sessions. Oral proficiency was measured at the beginning and end of the study. The findings showed that both the experimental and the control groups were able to significantly improve their oral proficiency from pretest to post-test, with the SCMC group showing a significantly higher development than the FTF group. Since the treatment of the experimental group included both SCMC and FTF interaction, it is unclear whether SCMC alone would lead to the same result or if learners do need actual FTF interaction to improve their oral skills. Nonetheless, the study demonstrated that SCMC enhances the development of oral proficiency beyond the level that FTF interaction alone could achieve.

Similarly, Blake (2009) investigated the effect of written SCMC on oral fluency in ESL learners. The participants followed the same six-week course in one of three conditions: the written SCMC group participated in sessions including chat group discussions, while the FTF group interacted with each other on the same activities in a FTF environment, and the control group did not interact with either the instructor or the other students during the course. Fluency gains were assessed based on pre-/post-test scores on five different fluency measures. The results revealed that the written

SCMC group showed significantly higher gain scores on two of the measures compared to the control group, while the FTF group's gain scores did not reach significance. Participants from the written SCMC group also significantly outperformed the FTF group. The study thus confirmed that learners can improve their oral fluency through written SCMC.

The following paragraph summarizes the current state of research on written SCMC focused on learners' performance and conducted within the cognitive-interactionist approach. The findings on learners' task performance indicate that written SCMC may lead to learner production with higher output and increased linguistic accuracy and complexity in terms of elaboration and variety of forms compared to FTF interaction, indicating that this mode provides learners with similar opportunities to produce output and develop their linguistic competence. Moreover, written SCMC may enable learners to produce pragmatic forms (e.g., address forms, discourse functions, speech acts) and benefit from interlocutor feedback on their use. Although the written SCMC may tend to generally lower learners' expectations of formality and appropriateness, it could be observed that the appropriate use of language can be instructed or negotiated in the interaction.

Research conducted from the cognitive-interactionist approach has produced evidence that written SCMC may provide learners with many of the same interactional features as FTF interaction, although SCMC is not as extensively researched as the FTF context (Ziegler & Mackey, 2017). The existing studies have shown that SCMC provides learners with opportunities to receive feedback that benefits learners to the same degree as feedback in FTF interaction. In qualitative terms, feedback in written SCMC appears to be noticed better by learners but tends to be less direct and immediate and involve less uptake and modified output compared to the feedback in FTF interaction. Learners have also been shown to engage in negotiation of meaning with an interlocutor through written SCMC. While negotiation of meaning in written SCMC seems to appear less frequently than in FTF interaction due to increased monitoring, longer processing time, and quicker turn-taking, the specific characteristics of the feature such as explicitness and increased noticing might positively impact its impact on the acquisition of L2/FL forms. Recent cognitive-interactionist research has also produced evidence for the occurrence of lexical and structural alignment in written

SCMC, with structural alignment being more frequent in written SCMC compared to FTF interaction given that the form is at the same time relatively difficult to produce, yet easy to type and familiar to learners. There is also some initial evidence that structural alignment may facilitate the development of L2/FL forms.

In terms of its impact on overall L2 and FL learning, written SCMC has a slight advantage compared to FTF interaction. On the question whether written SCMC may facilitate skill transfer and the development of oral skills, specifically, research to date has produced conflicting results with some studies indicating that written SCMC might have a small disadvantage compared to oral FTF interaction in facilitating the development of oral skills, and others suggesting that written SCMC may enhance the development of oral skills and may have an advantage to FTF interaction in promoting certain aspects of oral proficiency (e.g., fluency).

2.4 Dialog Systems for Language Learning

As explained in the previous chapter, L2 and FL interaction may be computer-mediated, enabling learners to engage in written or spoken interaction with an interlocutor through a computer. As a further form of computer-mediated interaction, learners may interact with the computer itself through a dialog system.

Learner interaction with a computer as a virtual language partner has been investigated by CALL research since the 1980s, using different terms to refer to the systems including, for example, “chatbots”, “robots”, “conversational agent”, or “intelligent personal assistants” (Bibauw, François, & Desmet, 2022, p. 121). The present thesis uses the umbrella term *dialog system* to refer to all applications allowing learners to engage in FL or L2 interaction with a computer through writing or speaking, optionally including non-verbal communication.

The heterogeneous use of terms reflect the variety of dialog systems that have been implemented and researched in L2 and FL learning: Dialog systems may run on different devices such as desktop or laptop computers, smartphones, smart speakers, or robots, using a voice or chat interface, which sometimes includes an animated character representing the virtual interlocutor – referred to as *embodied conversational agent* – or may be part of a computer game world or other simulated environment

(Bibauw et al., 2022). Accordingly, dialog systems for language learning may support input that is written (including written transcription of spoken input through speech-to-text technology), spoken, or multimodal (i.e., written, spoken, as well as visual, gestural, and haptic) (Bibauw et al., 2019).

Dialog systems also differ in terms of their underlying architecture and technologies implemented to process the user input and produce a corresponding output, which has implications for the type of interaction that they enable the user to have. While a detailed technical description is beyond the scope of this review, the following explanation provides a general overview and background, as the technological design of the systems also determines to some degree their affordance in terms of L2 and FL learning and teaching, as will be shown below.

The various approaches used in dialog system development are commonly categorized into two categories: rule-based and task-oriented approaches. Rule-based systems use a set of rules to match the user's input against a predefined pattern to determine the system's response. Rule-based systems allow for unstructured "open-ended" conversations (Bibauw et al., 2019, p. 18) or 'chat', which is the reason why these systems are also referred to as *chatbots* (Jurafsky & Martin, 2023). Task-oriented dialog systems on the other hand, typically use a set of natural language processing (NLP) techniques to extract the user's intentions from their input within a given domain (e.g., travel, restaurant) and determine the current state of the dialog (i.e., summarize the information and constraints that the user has given so far) to perform the relevant action (e.g., give a response, book a hotel room, place an order for a menu item) (Jurafsky & Martin, 2023). In the current dialog systems developed for research and commercial use, these NLP techniques rely on machine learning, where the system 'learns', i.e., develops its own algorithm, based on a corpus of language data (Jurafsky & Martin, 2023). The dialog system developed for the present study is of the task-oriented type and the documentation of its architecture (see section Dialog System) will further illustrate this type of dialog system.

Applied to L2 and FL learning, the two types of technological approaches to developing dialog systems afford learners an opportunity to engage in very different kinds of interactions, ranging from open-ended free conversations to guided dialogs in which learners select their response from a set of predefined utterances. Bibauw et al.

(2019) have established a typology of dialog systems for language learning, identifying four different categories along a continuum of level of constraint they place on the learners' input in terms of meaning and form:

- Narrative systems place the most constraint on both meaning and form, allowing learners to select from a list of preformulated utterances with different meanings in branching dialogs.
- Form-focused systems constrain the meaning and/or form of the learners' input explicitly, e.g., by instructing learners to produce an utterance with a given meaning or by completing a gap fill with the corresponding form.
- Goal-oriented systems use the task-oriented approach described above and constrain the meaning of the learner's input implicitly by embedding the interaction in a context and task. The form of the input may or may not be constrained, e.g., by prompting learners to use a certain target form in their production.
- Reactive systems include chatbots and other systems that place no constraints on meaning and form, allowing learners to engage in open-ended interactions.

From a cognitive-interactionist and task-based perspective, goal-oriented systems can be viewed as the most relevant system type as they most closely mimic the kind of interaction where FL/L2 learner and interlocutor work towards mutual understanding as they attempt to solve a given task, activating cognitive processes and communicative competences relevant for language learning. As the present study is interested in the use of dialog systems in the context of competence- and CLT-based EFL learning and teaching, it implemented a goal-oriented system allowing for task-based learner-system interaction, which will be described in more detail in the next chapter. The following review will therefore focus on research conducted on goal-oriented systems in the context of L2 and FL learning. The review presents the current evidence from the existing literature on dialog system-interaction for L2 and FL learning and identifies the knowledge gaps on the affordances of dialog systems for L2 and FL interaction and learning, laying the foundation for the present study. The

presentation of the empirical evidence is preceded by an overview of the claimed benefits of dialog system interaction for language learning.

2.4.1 Claimed Benefits

CALL researchers have argued from a cognitive-interactionist perspective that dialog system interaction may provide learners with similar opportunities to receive input, produce output, and engage in interactional features along with their beneficial functions for L2 learning as FTF or SCMC interaction (Bibauw, François, Van den Noortgate, et al., 2022). As stated above, the most promising type of dialog system from this perspective are goal-oriented dialog systems that engage learners in task-based interaction requiring learners to work towards mutual understanding by focusing on both meaning and form. However, it needs to be pointed out that technological constraints and limitations of current goal-oriented systems limit their interactivity in different ways. Thus, the systems' ability to, for example, negotiate meaning or provide corrective feedback may be limited. On the other hand, noticing might be enhanced compared to FTF interaction due to the written mode (or transcription of the oral input) (Bibauw et al., 2019). It has furthermore been suggested that alignment is higher in the learner-system context compared to human interaction (Rosenthal-von der Pütten et al., 2016; Sinclair et al., 2019). It is therefore an object of ongoing dialog system L2 and FL research to identify the specific ways in which goal-oriented dialog systems afford learners the same opportunities with respect to output, input, and interactional features as human interaction.

The other system types mentioned above such as narrative, form-focused, and reactive systems by design limit learners' overall focus on meaning and their ability to produce output and engage in interactional features such as feedback or negotiation of meaning. Their potential relates to the benefits of discrete components of interaction. Thus, interaction with these dialog systems may allow for pre-modifying the input according to the learners' proficiency level (Bibauw et al., 2019), it may push learners to produce language with the intended meaning (Bibauw et al., 2019), or promote the learners' noticing of gaps in their own production through monitoring or comparing it to the system's model production (H.-S. Kim et al., 2020; Sydorenko et al., 2019), which

can be reinforced by feedback-enabled systems (Sydorenko, 2015; Sydorenko et al., 2019).

CALL research has pointed to further advantages of dialog systems in general such as the fact that they are always available (Bibauw et al., 2019; W. Huang et al., 2022), allowing learners to engage in interaction independent from the availability of peers, teachers, or native speakers as for example in the context of distance learning (Kim et al., 2020) or self-learning (W. Huang et al., 2022) but also in classrooms with limited opportunities for interaction (Bibauw et al., 2019). Dialog systems also allow learners to repeat a task an unlimited number of times, promoting repeated practice and the automatization and the fluency, accuracy, and complexity of forms (Bibauw et al., 2019; Sydorenko et al., 2019). They may guide the learner to be exposed to and produce target forms that they would avoid in less guided human interaction (Divekar et al., 2021). Moreover, dialog systems might positively impact student motivation and their willingness to use the L2/FL (Divekar et al., 2021; Fryer et al., 2020), as they give learners the impression of an authentic native-speaking environment (W. Huang et al., 2022), they constitute an interesting new learning tool for learners (H.-S. Kim et al., 2020), and provide a low-anxiety learning environment (Bibauw et al., 2019).

CALL research has also highlighted the potential of future systems with further functionalities such as the possibility to implement optimal feedback options (Bibauw et al., 2019), to flexibly adapt their output to the learner's individual level proficiency (Bibauw et al., 2019; Bibauw, François, & Desmet, 2022; Divekar et al., 2021; Jiang, 2022) and provide motivational support (Bibauw et al., 2019), or to track and monitor learner activity and progress (Bibauw et al., 2019; Jiang, 2022).

2.4.2 Empirical Evidence

Empirical research on dialog systems for language learning has been conducted since the 1980s on different dialog system types (Bibauw, François, & Desmet, 2022), addressing a range of aspects from system description and technical evaluations, including the comparison of different design choices, to the measuring of user attitudes and the effectiveness of dialog systems in facilitating language development. According to the present study's aim, the following review of previous dialog system

literature focuses on research on written task-based interaction enabled by goal-oriented dialog systems, specifically, and studies that investigated learner-system interaction in terms of task performance (i.e., complexity, accuracy, fluency, communicative adequacy) or from the cognitive interactionist perspective in terms of input, output, and interactional opportunities.

The review first presents the research on the overall effectiveness of dialog systems for L2 development. This is followed by studies relating to aspects of task performance and, finally, a survey of cognitive-interactionist research on dialog system interaction.

A first meta-analysis by Bibauw, François, Van den Noortgate, et al. (2022) produced evidence for the positive effect of dialog systems on language learning and provided exploratory insight into the differential effects of the diverse system and interaction types. The study meta-analyzed 17 effectiveness studies published until 2018, measuring the effect of dialog system interaction on different outcome variables (e.g., proficiency, accuracy, complexity etc.). The pool of studies included all system types (i.e., form-focused, goal-oriented, reactive, narrative) with written or oral modality providing either no feedback or implicit or explicit feedback. The authors found an overall medium effect ($d = 0.59$) for dialog system interaction on development. With respect to the specific learning outcomes, the effects seem strongest for productive skills, overall proficiency, lexical and grammatical development, and accuracy, whereas the effect on fluency, complexity, and reading and listening comprehension could not be established due to insufficient sample size. The authors noted that with respect to overall learning outcomes, the overall effect size was within the range of the effects found in meta-analyses on both FTF human interaction and written SCMC, but the effects on specific dimensions of acquisition (e.g., grammar or vocabulary) were lower. According to the authors, this indicates that at the current state, dialog systems still have shortcomings which prevent them from providing learners with the same opportunities as human interaction does and need to be viewed as complementary rather than alternative to traditional interaction types. With regard to the effectiveness of moderating factors such as, for example, system type or experimental setting, the results are exploratory due to the low number of studies. Thus, both form-focused and goal-oriented systems appear to be beneficial for

language learning, while effects of narrative and reactive systems were inconclusive due to the low number of studies. With regard to the dialog type, system-guided interactions with high meaning constraints might lead to higher effects than task-oriented less constrained interactions, possibly due to enhanced conversation design in system-guided interaction which may, for example, introduce target structures and increase complexity progressively, or because the outcomes that were assessed in system-guided interaction were more achievable. Similarly, dialog system interaction with feedback, longer treatment (number of sessions, time on task), and high-intensity training led to larger effects than interaction with no feedback, shorter treatments, and spaced practice. Another noticeable observation was that lower proficiency learners (A₁ and A₂) might especially benefit from dialog system interaction, possibly because they work on consolidating newly learned linguistic knowledge in production, while the effects seem to diminish for more advanced learners (B₁ and B₂). It could further be found that written and spoken modalities led to almost identical effect sizes. Moreover, transfer from written practice to oral acquisition and, vice versa, from oral practice to written acquisition appeared to occur; however, it needs to be noticed that matching practice and test modalities led to much higher gains.

Additional recent studies confirm the effectiveness of dialog systems in facilitating language development, also including pragmatic development and providing additional insight into the factors moderating the learning effects such as modality or affective learner factors. For instance, Sydorenko et al. (2018) investigated the impact of dialog system interaction using video simulations on L2 pragmatic development focusing on requests. Twelve ESL learners with advanced proficiency participated in the study. Video simulations placed the participants in six variations of a scenario in which students requested a letter of recommendation from their instructor. Scenarios 1 and 2 required the participants to make requests in the role of students, scenarios 3 and 4 put them in the role of the instructor and thus enabled them to hear model utterances/requests, and scenarios 5 and 6 were a repetition of the first two scenarios. In each scenario, the learners first watched a video scene, then recorded their response, were taken to the next video scene corresponding to their specific response and so on. Changes in participants' oral production were measured in a pre-/post-test design with the first two simulations serving as pre- and the last two

simulations as post-tests. A microgenetic approach was used to analyze whether the participants' requests became more congruent with the system's pragmatically appropriate models in terms of form (e.g., use of modelled expressions such as "thank you for your consideration", "relevant information" etc.) and content (e.g., use of modelled request modifiers such as appealers, disarmers, grounders etc.). The findings showed that there were a considerable number of changes from pre- to post-test, with content changes being more frequent than form changes, which, according to the authors, might have been a result of learners' general focus on meaning over form in meaning-focused activities and the shortness of the intervention. Moreover, the occurrence of changes adopted from models was much more frequent than changes resulting from practicing, suggesting that while for some learners practicing alone was enough, model input played a bigger role in promoting request development. Generally, the simulations led to pragmatically more appropriate requests; however, in some instances the changes by some learners were not fully successful. This indicates that while the dialog system interaction did enable learners to notice and adopt expert speaker models, some learners, possibly especially those with less cultural knowledge and lower pragmatic and linguistic competence, might need additional practice opportunities and instruction. In sum, the study provided evidence for the beneficial effect of dialog systems for L2 pragmatic development and highlights the role of model input, at least for some learners.

Similarly, Alemi and Haeri (2020) investigated FL pragmatic development in dialog system interaction measured by the development of two speech acts (i.e., thanking and requests). The participants were elementary EFL learners at an Iranian kindergarten, with one group interacting with the robot acting as the 'teacher's assistant' and the second group interacting with the teacher alone for one hour on two days per week, for four weeks. The results from a post-test revealed that while both groups improved their production of speech acts, learners interacting with the robot significantly outperformed the control group. The authors ascribe the advantageous effect of the robot interaction on the learners' increased motivation and excitement as well as lower anxiety when interacting with the humanoid robot.

Kim et al. (2021) similarly compared the impact of written and spoken learner-system interaction as well as FTF peer interaction on learners' speaking performance.

In contrast to previous research, the study found an advantage for the voice-based dialog system over the other two interaction types. The study participants, Korean university English learners, were assigned to one of three groups: experimental group 1 who interacted with an unconstrained dialog system (the learners could choose between *Replika*, *Andy*, and *Google Assistant*) through written chat, experimental group 2 who interacted with one of the same three systems using the voice-chatting function, or the control group who engaged in learner-peer FTF interaction. The participants interacted with their interlocutor on the same conversational topics once a week during one semester. Changes in the speaking performance were identified by comparing the learners' mean scores in a pre- and post-test consisting of three speaking tasks. It was found that written as well as spoken learner-system interaction improved the performance in all three tasks, while FTF interactions improved the learners' performance in only two of the three tasks. According to the authors, learners interacting with the dialog systems were less anxious and more stimulated to produce significantly larger quantities of output than those interacting in FTF with a peer, which resulted in higher post-test gains for the dialog system group. This effect was strongest for one task, *describing a picture*, where the performance rating relied on the production of larger output quantities. A further finding was that, in one of the three tasks, *express an opinion*, the voice-based system group outperformed their pre-test performance to a higher degree than the other two groups. The authors suggest that the spoken dialog system mode might have encouraged learners to practice expressing their opinion more so than the other two interaction types but point out that the differential effect of spoken and written dialog systems need more research. In sum, the study confirms that both written and spoken dialog systems are effective in improving learners' speaking competence. In addition, the study provided evidence that dialog system interaction is equally effective as FTF interaction in promoting FL speaking and that it might be advantageous in promoting specific dimensions of learners' FL speaking competence, which might be related to their willingness to communicate. Similarly, the study points to the possibility that spoken dialog systems might have an advantage over written dialog systems in terms of specific aspects of FL speaking competence.

Kim et al. (2020) compared the impact of dialog system and FTF interaction on learners' writing performance as well as learners' attitudes and perspectives. University EFL learners completed a writing pre-test as well as a pre-survey measuring their attitudes toward learning English. Participants from the system group selected one of two unconstrained dialog systems with written modality, i.e., the chatbot *Replika* or *Andy*, and interacted with it on conversational topics for 10–15 minutes per week for one semester. Participants from the FTF group interacted orally with a randomly assigned peer on the same topics for the same duration. At the end of the semester both groups completed a post-test and post-survey assessing their writing performance and attitudes towards English learning, respectively. The results revealed that learners from both groups significantly improved their writing performance in the post-test. However, the dialog system group outperformed the FTF group in both writing tasks of the post-test. This indicates that written learner-system interaction compared to FTF spoken interaction provided additional benefits to learners for improving their writing competence. According to the authors, this might have been due to additional writing opportunities in dialog system interaction resulting from the written mode and overall increased production quantity to produce the foreign language in the system group. Lower anxiety and increased motivation in the dialog system group compared to the FTF group which learners indicated in the post-survey might also have positively influenced their performance in the post-test. So, the study showed that written dialog system interaction can also improve learners' FL writing skills and that its beneficial effect on FL writing performance might be greater than that of FTF interaction. The study furthermore suggests a positive impact of dialog system interaction on learner attitudes in terms of anxiety and motivation, which in turn might increase their FL performance.

However, studies have produced heterogeneous findings on the affective impact of dialog system interaction and its influence on performance. In line with Kim et al. (2020) reviewed above, many studies report positive learner reactions after interacting with dialog systems, such as high levels of motivation and engagement and low levels of language anxiety (Alemi et al., 2015; N.-Y. Kim, 2017a; Kohnke, 2023; Tai & Chen, 2023). Other studies instead found no change or an increase of anxiety level in learners after interacting with a dialog system (Çakmak, 2022; El Shazly, 2021), possibly because

learners might have had too little time to familiarize themselves with the dialog systems and experienced communication with them as more challenging than interacting with a human interlocutor who can adapt to the learners' needs, which also made learners more aware of their (perceived) linguistic inadequacies (El Shazly, 2021). Other factors that might impact learner motivation negatively are proficiency, with highly proficient learners being less motivated to interact with dialog systems, or general discomfort interacting with a non-human agent (Çakmak, 2022). Some studies found that the positive motivational effect might wear off as the novelty effect of interacting with a dialog system fades and, as learners improve their competence, they no longer find the interaction stimulating and challenging enough (Fryer et al., 2017; Sydorenko et al., 2019). Interestingly, however, these studies found that despite the unfavorable levels of anxiety and motivation, learners' performance nonetheless improved as a result from interacting with a dialog system (Çakmak, 2022; El Shazly, 2021). So, dialog system interaction appears to have a positive impact on affective factors in some learners but not others, with system familiarity and acceptance, ease of communication, and lower L2/FL proficiency possibly positively influencing the system's affective input on learners.

With regard to task performance in dialog system interaction, Nagata et al. (2020) explored the quantity and quality of learners' written production in dialog system interaction. In their study, college EFL learners interacted with an implementation of *ELIZA*, a written dialog system with no constraints on form, on two conversational topics. In the second condition, learners wrote a text on the same topics using a standard editor. The results showed that learners in the system group produced both more tokens and more types than the text editor group; however, the results were not statistically significant. A further finding was that learners in the system group used words unique to this condition. Further analysis revealed that learners had adopted these unique words from the system production, suggesting that the dialog system may introduce new vocabulary to learners. Similarly, learners in the system group had a lower error rate in the use of definite and indefinite articles, presumably because the dialog system had modeled their accurate use. Moreover, in instances where the system repeated the same question twice due to its technical shortcomings, the majority of learners improved the accuracy of their utterances when

answering the question again, although learners evaluated question repetition as a motivational drawback in post-questionnaires. In sum, this study provided empirical insight into different aspects of dialog system interaction: Firstly, it showed that system interaction may lead to at least the same quantity of text as well as variety of different words as non-interactive written production. Secondly, the study provided additional evidence for the occurrence and efficiency of the interactional features alignment and negotiation of meaning: The system was found to introduce novel vocabulary to the learners by modeling the accurate use of grammatical forms, i.e., through learner-to-system alignment, and assisted accuracy by pushing learners to produce modified output during question repetition, which can be considered instances of negotiation of meaning.

Focusing on the pragmatic appropriateness as an aspect of task performance, the study by Timpe-Laughlin and Dombi (2020) explored requests in learner-system interaction. Specifically, the authors investigated what request strategies and external and internal modifications FL learners used when interacting with a spoken dialog system. The learners were EFL speakers from two different universities in Japan and Hungary and with different L1s. They were asked to engage with the dialog system in a simulated video call with their supervisor in which they were required to make two requests. The requests were subsequently identified and coded for level of directness, request strategy, external modifiers, and internal modifiers. Overall, the majority of requests had a high level of directness and were realized by direct requests such as, e.g., "Uh I want you to have a meeting with me." (61.4%), followed by indirect requests with a medium level of directness using routinized formulae such as, e.g., "Can you revise my slide?" (37.6%), and hints with the lowest level of directness, e.g., "Uh I submitted a slide before attend the meeting." (1.1%). The authors hypothesized that hints, i.e., low level of directness, may have been rarely used as the learners considered them too subtle and reliant on context for the dialog system to understand. The most frequently used request strategies were want-statements such as, e.g., "Hello, I want to have a meeting with Lisa Green" (38.1%) and query preparatories, e.g., "Can I have a meeting with you?" (37.6%). The authors further observed a pattern of participants using more direct strategies with the first request they had to make compared to the second request, possibly because the first request was prompted by a direct question in

the dialog such as “What can I do for you?” and might also have been perceived as less imposing (requesting a meeting) than the second request (requesting a review of presentation slides). The general tendency for using more direct request strategies may, according to the authors, be explained with the participants’ low level of proficiency. External modification was used rarely (12.70% of requests), most frequently featuring grounders, imposition minimizers, and discourse orientation moves – depending on what was requested. The authors suggested that the low level of external modification might have been due to writing-economic reasons as well as to the automated interlocutor who reduced learners’ rapport building and negotiation efforts. Internal modification, on the other hand, i.e., the use of syntactic markers such as conditionals or lexical markers such as “please”, was quite frequent, with 36.5% of all requests containing one type of internal modifier and 20.1% of requests containing multiple internal modifiers. According to the authors this indicates the learners’ readiness to use highly routinized request forms that they knew from EFL textbooks. It was further observed that the L1 Japanese and L1 Hungarian speaker subgroups differed in the frequencies with which they produced direct and indirect requests, the type of request strategies and internal vs. external modifications used, possibly as a result of different English proficiency and L1 transfer. In sum, the study showed that FL learners, when interacting with a dialog system, use some of the same request strategies and modifications previously observed in L2 and FL FTF interaction. However, the learners’ realization of requests appeared to be, at least to some degree, influenced by the dialog system interaction type, even though these influences cannot be conclusively determined as the study lacked a comparison group: First, learners might have chosen more routinized, direct, and explicit requests as they expected more subtle requests to be missed by the non-human interlocutor. Second, the system’s modeling of requests appeared to trigger learner requests of a similar type. Third, learners seemed to be ready to use only highly routinized internal modifications rather than the more ‘uneconomical’ external modifications, possibly because they felt it was unnecessary to build rapport with a non-human agent.

The number of studies that explored dialog system interaction from a cognitive-interactionist perspective is still limited. The existing studies have focused on investigating whether and to what degree dialog system interaction may offer learners

with the same opportunities and benefits as human interaction in terms of receiving comprehensible input, producing output, and engaging in interactional features. Thus, a few studies have investigated the occurrence, nature, and acquisition benefits of interactional features such as feedback, negotiation of meaning, and alignment in the interaction of L2 and FL learners with goal-oriented dialog systems.

In a series of studies, Wilske and Wolska (Wilske, 2015; Wilske & Wolska, 2011; Wolska & Wilske, 2010a, 2010b) examined the effect of feedback and form-focused instruction in dialogue-based CALL on L2 development. In her doctoral thesis, Wilske (2015) reported on a study on the development of grammatical accuracy in the use of two L2 German target forms (dative case in prepositional phrases and word order in subordinate clauses). In the study, L2 German learners interacted with a dialogue system in written mode on two different communicative tasks (i.e., giving directions, making an appointment) under one of three conditions: unconstrained input with recast feedback (i.e., the learner chooses the linguistic realizations freely), unconstrained input with metalinguistic feedback, or constrained input (i.e., the learner is constrained to providing the target form in a pre-scripted gap-fill). Analysis of pre-post-test scores revealed that the constrained condition had more immediate effects on accuracy than the two unconstrained conditions, while the unconstrained conditions led to some immediate improvement. The author suggested that this was likely due to the learners avoiding the target structure in the unconstrained condition. In addition, the unconstrained recast group showed more delayed effects compared to the metalinguistic feedback group and the constrained group. According to the author, this result is in line with research on FTF interaction which showed that explicit instruction may lead to faster but less sustainable learning compared to implicit instructions. In sum, this study provides evidence on different aspects of task-based dialog system interaction: First of all, it provided evidence for the beneficial effect of task-based system interaction for the L2 grammatical development. Secondly, feedback types in system interaction impact L2 acquisition of forms in similar ways as in FTF interaction, with explicit feedback leading to immediate and implicit feedback to more sustainable acquisitional gains. Thirdly, both dialog systems constraining the form of the learner input as well as those with no constraints on form may promote L2

grammatical development, however, the constrained type appears advantageous, at least in terms of promoting short-term gains.

Kim (2017b) compared the occurrence and types of negotiation of meaning between spoken dialog system interaction and oral SCMC. The study participants, university English learners at a Korean university, of low, medium, and high proficiency interacted on a variety of conversational topics either in learner-peer dyads through SCMC or with a reactive, i.e., unconstrained, dialog system (i.e., 'chatbot') once a week over the course of ten weeks. The interaction transcripts were coded for number and type of negotiation, i.e., confirmation check, comprehension check, clarification request, repetition, reformulation. The results revealed that interaction with the dialog system led to more active negotiation of meaning compared to oral SCMC. More precisely, learners at the lower proficiency level used more repetitions and reformulations than those in the oral SCMC group. Learners at the medium level used more confirmation checks, clarification requests, and repetitions. High-proficiency learners used more confirmation checks and clarification requests. In addition, comparing the negotiation frequencies of the first and last chat showed that low-proficiency learners moved to using less repetitions and more reformulations over time, while medium-level learners used more clarification requests, and high-level learners used more confirmation checks. In sum, the study provided evidence that negotiation of meaning does occur in spoken dialog system interaction to at least the same extent as in oral learner-peer SCMC. Moreover, interaction with a dialog system might lead learners to increase their use of specific negotiation strategies over time, depending on their proficiency level.

Similarly, Satar and Qinghua (2020) explored the frequency and patterns of negotiation of meaning in the interaction between L2 learners and two different dialog systems. In the study, four lower- and four higher-proficiency Chinese EFL learners interacted with two different text-based reactive dialog systems allowing users to engage in conversational dialogue, *Mitsuku* and *Tutor Mike*, the latter of which additionally provided corrective feedback to the users. In the conversations, meaning was implicitly constrained by the task which required learners to chat about pre-set topics and questions. The chat logs were coded for negotiation of meaning routines, trigger types (i.e., lexis, syntax, discourse, content), and instances of modified output.

The findings indicated that both dialog systems led to negotiation of meaning in both groups. The lowest number was found in conversations of higher-level learners with *Tutor Mike*; however, the difference was not statistically significant, possibly due to the small sample size, according to the authors. The other three conditions led to similar frequencies. An analysis of the negotiation routines and components showed that all four conditions produced to varying degrees of incomplete negotiation routines where the routine did not include a response. However, the majority of negotiation sequences consisted of complete trigger-indicator-response-reaction to response routines, except in the high-proficiency group interacting with *Tutor Mike*, where 50% of negotiation sequences were abandoned after the indicator move. The authors point out that indicators sometimes did not signal non-understanding explicitly but were out-of-context and thus worked as “de-facto indicators” (Satar and Qinghua, 2020, p. 405). The triggers causing communication breakdowns varied depending on proficiency level and dialog system: In the interactions between low-proficiency learners and *Tutor Mike*, the most frequent triggers were lexical items (50%), followed by content (25%), discourse (17%), and syntax (8%). For high-proficiency learners the only triggers were discourse (57%) and content (43%) issues. Negotiation in the interactions with *Mitsuku* was most frequently triggered by content in both low-proficiency (43%) and high-proficiency (39%) groups. According to the authors, the high percentage of content-level triggers indicates the dialog system’s inability to conduct a meaningful conversation. In addition, both low- and high-proficiency learners produced modified output, i.e., semantic and morphosyntactic modifications of the trigger utterances during the response move, in 21% to 33% of all negotiation instances. The highest percentage of modified output was found in the low-proficiency group interacting with *Tutor Mike*, which indicates that the provision of feedback might facilitate the production of modified output in lower-proficiency learners. In sum, the study confirmed Kim’s (2017b) finding that dialog system interaction does lead to negotiation of meaning, with the frequency and characteristics varying depending on the learners’ proficiency level as well as the specific dialog system used: Negotiation of meaning in system interaction tends to generally be triggered by content and discourse issues, except in interaction between lower-proficiency learners and feedback-enabled system, where negotiation of meaning was most frequently triggered by lexical items.

Furthermore, lower-proficiency learners seem to produce more and more complete negotiation of meaning routines and feedback-enabled systems appear to lead to an increased ratio of modified output, the latter finding providing some initial evidence on the potential of negotiation of meaning in facilitating L2 acquisition.

Sinclair et al. (2019) examined alignment in ESL student-dialog system and student-tutor interaction. As learner-system corpus, the study used chat scripts from a commercial learning platform, on which L2 learners responded in written language to a reactive dialog system's questions on a variety of small-talk topics (e.g., "What sport do you play?" – "I play volleyball and I go running"). As human-human corpus the Barcelona English Language Corpus (BELC) was used, consisting of oral tutor-guided conversations with ESL learners on a range of conversational topics. Alignment in the two corpora was measured by identifying sequences of tokens ("expressions") that are used by both interlocutors. In learner-system interaction the most commonly aligned expressions involved the learners' re-formulation of the system's question (e.g., "What is your favorite day of the week?" – "My favorite day of the week is Friday") and the learner reflecting a question back to the system (e.g., "Where do you live?" – "I live in [location], where do you live?"). In learner-tutor interaction, on the other hand, alignment typically consisted of the tutor re-phrasing the learner's answer (e.g., "You like going out with your friends, good.") and the learner repeating the tutor's scaffolding moves. The degree of alignment in both interaction types was measured by "expression repetition", i.e., the ratio of expressions to total tokens produced. The results revealed significant learner-system alignment; however, no system-learner alignment was found, which the authors explained with the fact that the system was not designed specifically to align to the user. Furthermore, learner alignment was weaker in learner-system interaction than human-human learner-tutor interaction. According to the authors this might have been due to the lack of learner support that was found in the human-human corpus. Another finding was that in both interaction types, learners aligned significantly more to their interlocutor than vice versa. Analyzing the nature of the alignment, learner-system interaction produced a greater variability of alignment compared to learner-tutor interaction, with low-level alignment indicating lack of student engagement and high-level alignment possibly indicating the learners' copy-pasting or continual repetition of the system's utterances.

Despite the study's limitation of having used corpora that differed in multiple aspects such as modality and learner L1 and proficiency, it provides some evidence for the existence of lexical alignment in learner-system interaction. However, in systems not specifically designed to align, this finding seems to only apply to the direction learner-to-system and not vice versa. Furthermore, alignment from learner to system was weaker than from learner to human tutor, as learners experienced less support by the dialog system. Low levels of student engagement also appear to decrease learner-to-system alignment, whereas specific student strategies such as copy-pasting of previous system utterances increased the alignment rate.

Rosenthal et al. (2016) examined whether dialog system embodiment and quality of speech output had an impact on the learners' alignment to their virtual interlocutor. The participants, adult L2 German learners, interacted with a dialog system under one of six conditions: with one of three types of dialog systems (language-only, virtual version of the *Nao* robot, physical *Nao* robot) with text-to-speech generated speech output or with prerecorded natural speech. To measure alignment, the learners interacted with the assigned dialog system on two question-answer tasks. The learners' utterances were subsequently analyzed for lexical alignment (i.e., ratio of reused lexical item over occurrences of the concept) and structural alignment (i.e., ratio of reused grammatical case per every five sentences). The results revealed that while learner-system alignment occurred in all conditions neither embodiment nor the quality of the speech output had an effect on the degree of alignment. Pre- and post-measurement of the participants' language performance furthermore showed that the performance deteriorated in the post-test compared to the pre-test. The authors suggested that this was due to the high cognitive effort involved in the post-test, which participants completed after the 50min interaction with the dialog system. Of the three system types, the physical robot group performed worse in the post-test compared to the other two groups, possibly because they were more focused on the robot than the task itself during interaction, according to the authors. The participants' alignment ratios during interaction had no significant effect on their performance in the post-test, possibly due to the short treatment time and a lack of proficiency in some learners which might have prevented them from benefitting from the alignment. In sum, the study provided further evidence for the

occurrence of lexical learner-system alignment as well as initial evidence for the occurrence of syntactic learner-system alignment, which was not impacted by dialog system embodiment and type of speech output. However, due to the above-mentioned limitations, the study could not provide any evidence for the positive effect of learner-system alignment on L2 performance.

To summarize the current state of research on dialog systems for language learning, there is emergent evidence for the effectiveness of dialog system interaction for language development in terms of overall learning, with an effect size that is slightly lower than that measured in human interaction contexts (e.g., FTF, SCMC). While evidence on the differential effects of the different system types is still outstanding, goal-oriented systems, specifically, appear to have an effect on their own. With regard to the different modalities, written and spoken modalities appear to be equally beneficial for language learning and there seems to be some transfer from written practice to oral acquisition and vice versa.

In terms of the specific types of learning outcomes, there is initial evidence that current dialog system interaction is beneficial – and potentially more beneficial compared to human interaction – for the development of certain aspects of learning (i.e., productive skills, lexical and grammatical development, general proficiency, accuracy, pragmatic forms), while effects on other dimensions (e.g., complexity or fluency) remain to be demonstrated. However, current dialog systems seem to still be disadvantageous compared to human interaction in promoting more focused dimensions of acquisition, e.g., specific lexical, grammatical, or pragmatic features. Possible advantages of dialog system over human interaction related to certain aspects, as mentioned above could be a result of higher production quantity and more positive affective learner states in dialog system interaction. However, learners appear to only feel more positive about dialog system interaction if they are familiar with and accepting of the system and have lower proficiency. Learners with lower proficiency seem to generally benefit more from dialog system interaction than more advanced learners.

With regard to interactional features, initial evidence suggests that they share some characteristics with those occurring in human interaction: For instance, as in

human interaction, the presence of feedback in dialog system interaction seems to positively impact L2/FL development, with more explicit feedback leading to short-term, and implicit feedback leading to long-term gains. It has also been demonstrated that there is negotiation of meaning in system interaction, occurring with at least the same frequency as in human interaction. However, dialog system interaction does seem to influence the nature of negotiation of meaning, even if the nature of meaning negotiation varies from system to system: Thus, learners have been shown to change their negotiation behavior after long-term use of dialog systems. Negotiation of meaning in system interaction also tends to be frequently triggered by content issues as a result of the systems' lack of understanding and are often incomplete, i.e., lack a response move. It remains to be demonstrated if and how these characteristics of learner-system negotiation of meaning impacts the effect of the interactional feature on the acquisition of L2/FL forms. Similarly, dialog system interaction may lead to lexical and structural alignment, however, there is currently only evidence for alignment in the direction from learner to dialog system and no evidence for its effect.

Research on learners' performance in terms of complexity, accuracy, and fluency as well as communicative adequacy when engaging with a dialog system in task-based interaction is still very limited. Existing studies suggest that dialog system interaction could enhance complexity and accuracy as learners benefit from the system's modeling. However, learners' pragmatic appropriateness could be negatively affected when interacting with a non-human agent, however, higher learner proficiency and the modeling of appropriate forms might counteract this tendency.

2.5 Research Gaps and Derivation of Research Questions

As stated in the introduction, this study aims to explore the potential of task-based interaction between EFL learners and a dialog system for commercial VET FL classroom learning and teaching. Following the review of the literature on the topic in the last section, this section identifies the gaps in the prior research and presents the research questions developed for the present study to address these gaps.

As argued above, the cognitive-interactionist and task-based approaches constitute a suitable framework to investigate dialog system interaction in the

commercial VET EFL learning and teaching context: As became evident in the review of the existing literature, the cognitive-interactionist approach has well-established theoretical concepts and a substantial body of empirical research providing evidence for its claims. Thus, cognitive-interactionist research interaction has shown that the learners' opportunities during interaction to receive comprehensible input, produce output, and engage in interactional features to facilitate language learning. Substantial evidence for these relationships has been found for FTF interaction and, to a lesser degree, for SCMC interaction.

The observation that the concepts and claims of the cognitive-interactionist approach also apply to the SCMC context underpins the rationale to investigate the modally similar learner-system interaction from a cognitive-interactionist perspective. While CALL research has begun this endeavor, more evidence is needed to determine to what degree learner-system interaction provides learners with the same opportunities for input, output, and interactional features as FTF and SCMC and whether these components have similar acquisition effects. The present study will help to investigate further how far learners may benefit from interacting with a dialog system to develop their FL.

Cognitive-interactionist research on dialog system interaction has so far been able to show that dialog system interaction may promote language development and that the effect could, in fact, be greater with respect to some aspects of language learning compared to human interaction. Research on L2 and FL dialog system interaction from the cognitive-interactionist perspective has also identified the occurrence of initial interactional features, i.e., negotiation of meaning and alignment, and begun to describe their interaction-specific characteristics. The few existing studies indicate both similarities and differences between the features occurring in learner-system interaction and those occurring between learner and a human interlocutor. The interactional feature *feedback* has so far only been explored in learner interaction with form-focused dialog systems specifically designed for the purpose but not in task-based interaction with less constrained goal-oriented systems.

A second important aspect in assessing the potential of learner-system interaction in the commercial VET EFL context is the focus on task-based interaction. As argued above, CLT-based FL classrooms in commercial VET and other school

contexts rely on interaction activities that are based on tasks to promote the development of communicative competences (i.e., linguistic, pragmatic, sociolinguistic competences) in line with the CLT approach. Therefore, an evaluation of dialog systems for commercial VET EFL classrooms must focus on their affordance to engage learners in task-based interaction allowing them to produce language with sufficient levels of complexity, accuracy, fluency, and communicative adequacy to stimulate the development of the linguistic as well as pragmatic and sociolinguistic components of communicative competence. As becomes clear from the review on FL and L2 dialog system interaction, very few studies have assessed learner-system interaction in terms of task performance. The limited evidence produced by these studies points to a possible positive effect of system interaction on linguistic complexity and accuracy, while the evidence on its effect on pragmatic appropriateness, specifically, is mixed. Overall, more evidence is needed to determine how learner-system interaction compares to human interaction in terms of task performance and thus its affordances to engage learners in task-based interaction and stimulate the development of communicative competences.

To address these research gaps, the following research questions were developed:

1. *What interactional features occur in written task-based interaction between commercial VET EFL learners and a dialog system and what are their characteristics compared to those found in learner-peer written SCMC on the same task?*
2. *Is there a difference in the learners' task performance in terms of linguistic complexity and accuracy as well as communicative adequacy between learner-system interaction and learner-peer SCMC?*

In answering research question 1, the present study will contribute to a clearer picture of the nature of learner-system interaction, specifically, with respect to its affordances to provide learners with the opportunity to engage in the interactional features relevant for acquisition. Comparing the interactional features in learner-system interaction with those found in learner-peer SCMC on the same task will help

to identify, by way of contrasting, the specific characteristic of interactional features pertinent to learner-system interaction.

Research question 2 aims at contributing to the understanding of the potential of dialog systems for task-based interaction by assessing learners' task performance when interacting with a dialog system. Since the present study investigates written learner-system interaction, it excludes fluency as a dimension of task performance. Benchmarking the learners' task performance during system interaction against that of learner-peer SCMC, allows for the identification of the ways in which dialog system interaction may impact the complexity, accuracy, and communicative adequacy of learners' FL production in this type of interaction. Task performance is explored using a non-directional research question, as previous research on the topic has produced inconclusive evidence on the effect of dialog systems on task performance.

A dialog system with written rather than spoken or multimodal (i.e., written, spoken, as well as visual, gestural, and haptic) mode was evaluated for pedagogical as well as research considerations: Written chat constitutes an unintrusive technology and is a familiar mode of communication for young Swiss learners, which makes it well-suited for VET classroom settings. At the same time, it allowed for the collection of reliable data that is unimpacted by potential technological obstructions such as, for example, speech-to-text inaccuracies or learner difficulties when handling the speech-to-text function.

The exploratory and descriptive nature of the research questions is warranted at the current limited state of knowledge on these aspects of dialog system interaction. By describing in detail the occurrence and nature of interactional features as well as learners' task performance in terms of complexity, accuracy, and communicative adequacy in learner-system interaction, the current study establishes a foundation for further investigations of these aspects. For instance, future studies may include experimental studies on the effectiveness of interactional features and task performance for the development of FL/L2 competence, or studies searching to explain the advantages and disadvantages of dialog system interaction for promoting language learning.

From the perspective of FL and L2 teaching, the findings on the above questions may inform teachers on the affordances and limitations of goal-oriented dialog systems to engage learners in learning-facilitating interaction.

3 Methods

To review, the purpose of this study was to describe the nature of task-based interaction between EFL learners and a dialog system with the goal to further explore its potential for commercial VET EFL classroom learning and teaching. The research questions are stated again below.

1. *What interactional features occur in written task-based interaction between commercial VET EFL learners and a dialog system and what are their characteristics compared to those found in learner-peer written SCMC on the same task?*
2. *Is there a difference in the learners' task performance in terms of linguistic complexity and accuracy as well as communicative adequacy between learner-system interaction and learner-peer SCMC?*

This chapter describes the design and methods used to address these research questions. It first presents an overview of the research design and argues for the choice of the mixed methods approach used for the study. It then reports on pilot study 1 which served to develop the task that was used for data collection. The following section provides a brief description of the development and functions of the dialog system that was used to carry out the task in learner-system interaction. A rough understanding of the system's architecture and functions is necessary to be able to analyze and interpret the findings on the learner-system interactions. The subsequent two sections report on pilot studies 2 and 3 which served to test the infrastructure, materials, and procedure for the main study. This is followed by a report on the data collection of the main study (participants, infrastructure, materials, and procedure). The final section describes the procedures that were followed to process and analyze the data.

3.1 Research Design

The design chosen to explore the research questions is visualized in Table 1 and described in the text below. A detailed description of the instruments, data collection procedure, and data analysis will be provided in the following sections.

Table 1

Overview of Research Design for Main Study

Data collection	Data sources	Data analysis
Session 1: study introduction and language test	UCLES 2001 Quick Placement Test	Marking and calculation of test scores
Session 2: interaction and pre- and post-questionnaires	Scripts of learner-system interaction and learner-peer SCMC	RQ 1: Coding of interactional features; counting of codes and descriptive statistical analysis (proportions, frequency measures) RQ 2: Preliminary coding of units of analysis; calculating measures of complexity, accuracy, communicative adequacy; descriptive statistical analysis (proportions, frequency measures) and statistical analysis to test for differences between interaction types (independent sample two-tailed <i>t</i> -tests, Mann-Whitney <i>U</i> tests)
	Questionnaire reports	Preliminary grouping and counting of open format questions; descriptive statistical analysis (proportions)
Session 3: retrospective interviews	Interview transcriptions	Content-structuring qualitative content analysis

As shown in Table 1, data collection was conducted in three sessions to obtain the necessary data for the study. The instruments, infrastructure, and procedures used for data collection were previously piloted in three separate pilot studies which are documented in the first few subsections of the current chapter. During session 1, the study was introduced to the participants and they completed a language test. The test scores served to describe the participants and assign them to an interlocutor with

equal language proficiency when carrying out the interactive task during session 2 in learner-peer SCMC.

In session 2, learners carried out an interactive task either in learner-system interaction or learner-peer SCMC to generate the chat scripts used as main data in the study. To enable the learner-system interactions, a written goal-oriented dialog system was developed for the purpose of the study since no available dialog system was suitable to use in a curriculum-based task, as will be described in more detail below. The implemented task was based on the commercial VET EFL curriculum. To generate the learner-peer SCMC data, the SCMC chat application *Mattermost* (Mattermost, 2019) was used to carry out the same task with a second group of learners. Before and after carrying out the task in session 2, the participants completed a pre-questionnaire to provide their relevant previous experience and a post-questionnaire to report their perspective on the task.

In session 3, retrospective interviews were conducted with a selection of participants during which the learners self-reported on their thought processes, strategies, and emotional responses when interacting with the dialog system or a peer, respectively. Along with the post-questionnaire data, the interview data served to complement the finding and interpretation of the main data (i.e., interaction scripts) research.

As a first step in analyzing the main data (i.e., the learner-system and learner-peer chat scripts), the learner-system interaction scripts were coded for interactional features to address research question 2. To identify the specific nature of learner-system interactional features, each interactional feature was assigned subcodes pertaining to their unique characteristics. The same codes were subsequently applied to the learner-peer SCMC scripts and adapted where necessary, to highlight differences and overlaps between the two interaction types. The codes were also counted and measured for frequencies to identify patterns and trends in the occurrence and nature of interactional features in the two interaction types. To address research question 2, the scripts were initially coded for units of analysis used for calculating measures relating to linguistic complexity, accuracy, and communicative adequacy. The

calculated measures were then used to statistically test for differences between the two interaction types.

The questionnaire responses were analyzed by grouping the responses to the open-format questions and counting the frequencies of all responses to identify the relevant aspects and possible patterns and relationships between the different aspects. The interview data was analyzed by conducting content-structuring qualitative content analysis. The resulting codes were not quantified as the goal of the analysis was to capture the breadth of the learners' perceptions and reactions as they occurred in the two groups (system, SCMC) rather than to measure and compare the prevalence of a given phenomenon.

3.2 Description and Rationale for Use of Mixed Methods Approach

As becomes evident from the descriptions above, this study used a mixed-methods approach, which is defined by Kuckartz (2014), subsuming the different existing definitions of mixed methods, as the "Kombination und Integration von qualitativen und quantitativen Methoden im Rahmen des gleichen Forschungsprojekts" [combination and integration of qualitative and quantitative methods within the same research project] (p. 33): In a mixed methods design, researchers collect qualitative as well as quantitative data, integrating the data, findings, and conclusions from both research strands in the final phase or an earlier phase of the project.

The following paragraphs highlight the mixing and integrating of qualitative and quantitative methods at the different stages of the research design used in the present study and discuss the suitability and consequences of choosing a combination of methods: The two research questions proposed by this study involve qualitative as well as quantitative aspects, suggesting a combination of both qualitative and quantitative methods in the overall study design. While research question 1 has a qualitative focus, investigating the characteristics of interactional features, research question 2 focuses on the quantitatively measured constructs complexity, accuracy, and communicative adequacy.

Due to research-economic reasons, the same sets of learner-system and SCMC scripts had to be used to investigate both research questions, which implied that the sampling method had to agree with both a qualitative and quantitative approach: In line with qualitative sampling, the sample size, i.e., the number of interactions, was determined following the principle of saturation (Kuckartz, 2014). At the same time, the sample needed to agree with quantitative sampling principles such as statistical power and randomized group assignment. These considerations impacted the sampling, which will be described in more detail in the next section.

The collected interaction scripts were subsequently analyzed in two separate strands to address each research question individually, i.e., qualitatively analyzed for interactional features (research question 1) and quantitatively analyzed for performance measures (research question 2). However, at the analysis stage of research question 1, the codes were counted and measured for frequency, a “quantification” process which is considered another form of mixing qualitative and quantitative methods used to enhance the qualitative findings (Creswell & Clark, 2011, p. 81; Kuckartz, 2014, p. 87). As mentioned above, this helped to describe the characteristics of interactional features in the two different interaction types more completely.

Similarly, qualitative as well as quantitative methods were used when collecting and analyzing the additional data from learners’ questionnaires and interviews. The questionnaire, consisting of mainly close-ended questions with frequency scale response types as well as some open-ended format questions, produced quantitative as well as qualitative data. However, the focus of the questionnaire was a qualitative one, aiming to identify the learners’ thought processes and strategies. When analyzing it, the responses were therefore analyzed qualitatively by identifying and grouping themes and their relationships, additionally counting frequencies to find patterns. Similarly, the qualitative interviews served to understand what the learners’ thought processes, strategies, and emotional responses were when interacting. It was therefore analyzed qualitatively using content analysis.

The findings on the complementary questionnaire and interview data were used to help interpret the findings on the interaction scripts in both the qualitative strand relating to research question 1, and the quantitative strand relating to research question 2. More precisely, linking the learners’ self-reported thought processes,

strategies, and emotional responses to the findings on interactional features and performance aspects allowed for the tying together and interpretation of these observations and highlighted issues relevant for further research.

3.3 Pilot Study 1: Task Development

The purpose of pilot study 1 was to test and evaluate two alternative EFL tasks to be used in the main study.

The two tasks to be evaluated were designed based on the participants' curriculum and coursebooks in line with the study's purpose of investigating task-based interaction in the commercial VET context. This approach also guaranteed that the participants were familiar with the task content and equipped with the necessary linguistic and academic knowledge to carry out the tasks. As explained in the introduction, the commercial VET EFL curriculum is developed at the federal level and implemented in all Swiss VET commercial schools. The curriculum that was in place at the time of the present study provided a total of 240 lessons over two or three years (SKKAB, 2020). The students at the school where data collection was conducted followed three lessons of English per week and completed the subject at the end of their second school year. According to the curriculum, students were required to achieve B1 level according to the CEFR in the four language areas (listening, speaking, reading, and writing).

At this school, the federal curriculum had been worked out into a more detailed curriculum that was followed by all English faculty staff members (see Appendix A). When developing the task, the topics and objectives of the school curriculum were used as a primary foundation. Since the study participants were in their second semester at the time of data collection, it was ensured that the task included topics and objectives from the first and second semester only. These included a wide variety of both general and business English topics including, for instance, "request and give information, describe people and places, justify and exchange opinions, role plays, give and receive directions, exchange future events and possibilities, introduce yourself and receive guests" (for the full list, see Appendix A).

The task design was further based on considerations from task-based interactional research, specifically, the task features most likely to elicit interactional work such as modifications and the use of communicative strategies, as identified by Ellis (2003) (e.g., required information exchange, two-way information gap, closed outcome, no contextual support, need for detailed information).

The communicative limitations of the dialog system that would subsequently be developed and used for the study had to also be considered when developing the task. As will be seen below, dialog systems, at that point of technological development, had certain communicative restrictions restricting their use to certain dialog types. For instance, two-way information gap activities or collaborative dialogs were outside the scope of current dialog systems.

Based on these considerations, the following two tasks were developed. The task instructions were given in German. The complete handout containing the final task can be found in the appendix (see Appendix C).

- **Internship task:** In a role play between a commercial apprentice interested in an internship program abroad and the internship program coordinator, the apprentice's task is to ask the coordinator questions about three different internship opportunities based on a given detailed set of interests and qualifications in order to decide which is the most suitable. The coordinator holds the information about the three opportunities.
- **Hotel task:** In a role play between a commercial apprentice and a hotel employee, the apprentice is given the task of finding the most suitable out of three hotel event rooms, along with a list of detailed requirements. The hotel employee answers the apprentice's questions about the three rooms.

The tasks were tested in a class of 13 commercial VET students (nine female, four male) at the same school at which the main study took place. Since the tasks consisted of role-play activities requiring an even number of interlocutors, the teacher also participated in the interactions. The participants were informed about the procedure of the test and subsequently randomly assigned into seven dyads. Three dyads were given the internship task and four dyads were given the hotel task. The participants carried out the role play in their assigned role (i.e., customer or internship

coordinator in the internship task or customer or hotel employee in the hotel task), interacting with each other through the chat application *WhatsApp* on their personal smartphones. After completing the task, one person of each dyad emailed the exported chat log files to the researcher. In the second lesson, the dyads in each group were randomized again and assigned to the other task, while also reversing the task roles. Time was limited to a maximum of 20 minutes per task.

Analyzing the resulting total 14 chat scripts for interactional features and general performance (i.e., text amount) revealed that both tasks produced a comparable number of negotiation of meaning episodes (internship task: five episodes; hotel task: four episodes) and self-corrections (internship task: 15 self-corrections; hotel task: nine self-corrections). However, the hotel task resulted in a higher word count per interaction ($M = 554.3$) than the internship task ($M = 499.9$) and more task completions (hotel: four; internship: one). In addition, students and teachers expressed a preference for the hotel task over the internship task in their feedback following the chat sessions. Based on these observations, the hotel task was selected as a basis for the dialog system development and for use in the main study.

3.4 Dialog System

As mentioned above, at the time of the present study, no dialog system was available that would enable EFL learner interaction on a task that is relevant for the commercial VET curriculum. Dialog systems developed for commercial use outside of EFL learning contexts such as ‘chatbots’ built into vendor websites or social media platforms or virtual ‘personal assistants’ built into smartphones or smart speakers (e.g., Amazon’s *Echo*, Apple’s *Siri*, or Google’s *Assistant*) may offer learners the opportunity to practice the FL in free conversation (Fryer et al., 2020; Moussalli & Cardoso, 2016) but are not suited for interaction on specific curriculum-based tasks (Bibauw et al., 2019). The same applies to freely available unconstrained systems, i.e., ‘chatbots’, such as *Tutor Mike* (R. Lee, n.d.) or *Cleverbot* (Carpenter, n.d.), which are unable to carry out a conversation on a specific task. Of the systems developed specifically for language learning practice and research, only a few are publicly accessible (Bibauw et

al., 2019) and none of them enable interaction on a task that would be relevant for the commercial VET context.

It was therefore decided to develop a dialog system specifically for the purpose of this study, as it would allow for the implementation of a curriculum-based task in line with the study's purpose and research questions. As a limitation to this approach, using an own system would limit the generalizability and comparability of the findings, as each system can be expected to impact learner interaction differently to some degree, including the occurrence and nature of interactional features, task performance, and its benefits for language learning (Bibauw et al. 2022). To facilitate the comparability of the findings to those generated by future other studies, care was taken to use freely available technologies, as far as possible, and to document the system development transparently.

The dialog system was developed in collaboration with two bachelor students at the School of Engineering at University of Applied Sciences and Arts Northwestern Switzerland as part of their bachelor's thesis. The development took place over the course of the Fall/Winter semester 2018/2019. The researcher's role in the development process was to consult from an EFL learning and teaching point of view, specifically, to develop the task that was used as a basis for the dialog design, to contribute to the data generation for the system's training (see below), and to edit the dialog system's set of responses (see below).

The design requirements were to develop a dialog system that would allow its users to interact on the hotel event room reservation task described in the previous section, with the system taking on the role of the service provider as 'hotel chatbot'. In other words, the purpose of the dialog system was neither to create the illusion of interacting with a human interlocutor nor to mimic human interaction as closely as possible. Instead, the dialog system was to simulate human interaction closely enough for learners to relate it to the communicative context of the task. Following the typology offered by Bibauw et al. (2019), such a system can be described as a goal-oriented dialog system with contextualized dialog that places no constraints on form but constrains meaning implicitly by enabling interaction on the topic of the task only.

Further requirements from a research-methodological perspective were that the system should be browser-based for ease of use, enable multiple learners (i.e., up to a

complete class of 24 learners) to use the system simultaneously, assign a unique identification to each learner, and allow for the export of the chat log. Finally, as mentioned above, it was desirable to build the dialog system using publicly available state-of-the-art technology for reasons of comparability and generalizability.

The following section presents a short description of the dialog system that was developed and implemented in the present study. The aim of this description is to create an understanding of the system's architecture and functions as relevant for the analysis and interpretations of the learner-system interactions. A more detailed description and full report of the development process may be found in Louis and Cocquio (2019).

3.4.1 Overview

In line with the requirements stated above, the dialog system was built with an interface that was a web application client allowing learners to interact with the dialog system through the browser. The client also enforced the learners' authentication through entering a unique participant ID before being able to interact with the dialog system, provided the export of the chat scripts, and handled part of the error detection as will be described below.

Responding to the generalizability and comparability requirements, the dialog system was built using the open-source machine learning framework Rasa (Rasa Technologies GmbH, n.d.) based on its satisfactory classification ability on a set of evaluation data as well as additional criteria such as source availability, cost, data processing location, and customization (Louis & Cocquio, 2019). Dialog system frameworks offer dialog system developers the natural language processing (NLP) components, such as natural language understanding (NLU) components and the dialog manager necessary to build a dialog system, thus making the development process more efficient and accessible. While the NLU components are required to extract meaning from the user's input, the dialog manager controls the dialog as a whole by determining the state of the dialog and selecting an appropriate system response.

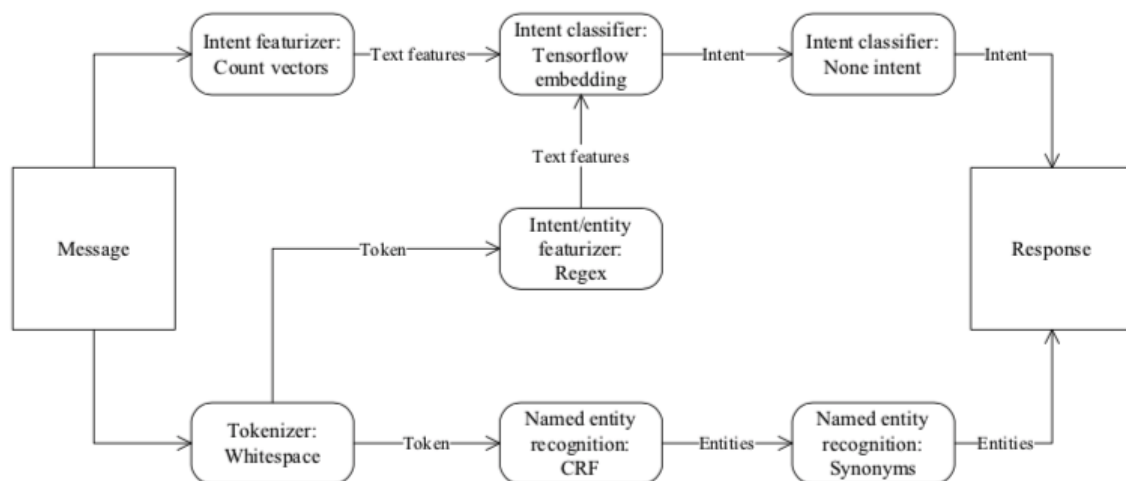
3.4.2 NLU

The framework Rasa used for the development in the present study offered a set of 20 natural language understanding (NLU) components that could be combined into an NLU pipeline as well as an application programming interface (API) to build custom components.

The dialog system developed for this project was built using a customized NLU pipeline, consisting of the components visualized in Figure 1. In its entirety, the NLU pipeline processes the user message, leading to the classification of the user's dialog acts, i.e., intents, and extracting of core information, i.e., entities.

Figure 1

Visualization of the Dialog System's NLU Pipeline



Note. From “Visualized Data Flow Between NLU Components in the Pipeline” by Louis and Cocquio, 2019, *Chatbot in the English Classroom. Encourage Negotiations of Meaning* [Bachelor’s thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland], p. 21. FHNW.

https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/thesis-cocquio_louis.pdf

Figure 2*Structured Representation of User Message by Dialog System*

```

{
  "intent": {
    "name": "greet+provide_name",
    "confidence": 0.9047530889511108
  },
  "entities": [
    {
      "start": 12,
      "end": 17,
      "value": "David",
      "entity": "name",
      "confidence": 0.9912837177405575,
      "extractor": "ner_crf"
    }
  ],
  "intent_ranking": [
    {
      "name": "greet+provide_name",
      "confidence": 0.9047530889511108
    },
    {
      "name": "provide_name",
      "confidence": 0.2864721417427063
    },
    ...
  ],
  "text": "Hello, I am David"
}

```

Note. From “An Abridged Response Object of Rasa NLU” by Louis and Cocquio, 2019, *Chatbot in the English Classroom. Encourage Negotiations of Meaning* [Bachelor’s thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland], p. 22. FHNW.

https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/thesis-cocquio_louis.pdf

The intent classifier determines the intent of the user messages. It is trained on training text data that is prepared (tokenized, count distinct words, encoded) by the intent featurizer. During training, the intent classifier compares the user messages with the associated intent in the labeled training data and builds a model that classifies new messages into the different intents. For the development of the current dialog system, 29 different intents were determined. If the intent classifier fails to map a message to an intent, the custom *None* intent classifier attempts to label it based on the entities found in the message.

The named entity recognition extracts information from a predefined list of semantic categories/objects (i.e., *entities*) in the user messages, e.g., the user’s name,

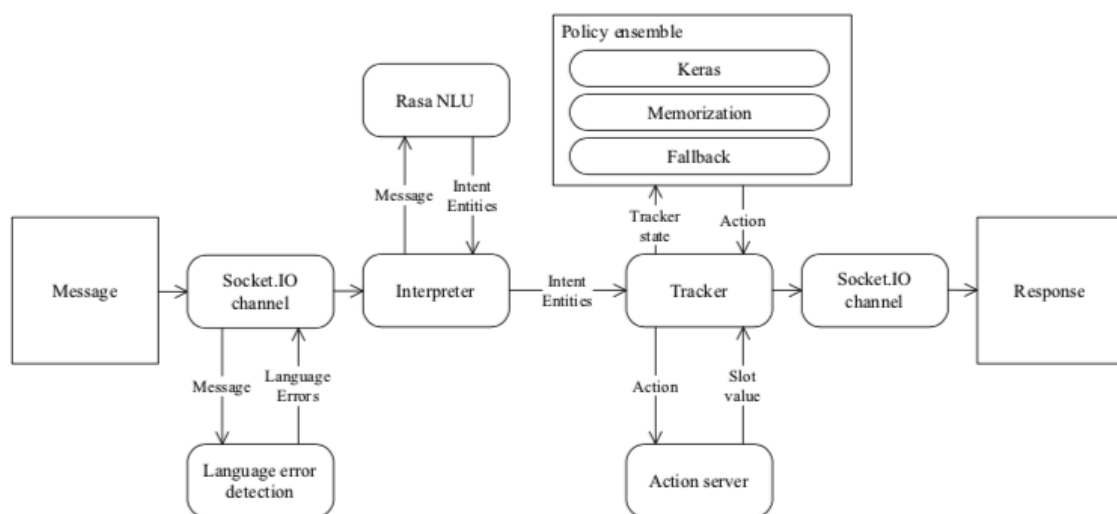
available budget, date of booking etc. The named entity recognition is trained on the training data, where all possible, i.e., featured, entities are labeled, calculating the probability of a given sequence containing a certain entity, e.g., entity *name* in “I am David”. At the end of this process, the NLU returns a structured representation of the user message (see Figure 2), consisting of the top-ranking intent, detected entities, all other intents with a confidence above zero and the original supplied message.

3.4.3 Dialog Manager

Rasa offers a dialog manager branded *Rasa Core*. The dialog manager implemented in the dialog system developed for the present study consisted of the customized pipeline of interacting components visualized in Figure 3.

Figure 3

Visualization of the Dialog System’s Dialog Manager Pipeline



Note. From “Visualized Data Flow Between Rasa Core Components” by Louis and Cocquio, 2019, *Chatbot in the English Classroom. Encourage Negotiations of Meaning* [Bachelor’s thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland], p. 22. FHNW.

https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/thesis-cocquio_louis.pdf

The message from the NLU is sent via the interpreter to the dialog manager. The first component of the dialog manager was the language error detection which was added to the standard dialog manager. Since the language ‘understanding’ of dialog systems relies on simplified model of the actual human understanding process, it may lead to very different results when interpreting learner input and might accidentally understand language that is highly erroneous or even nonsensical. To prevent this and encourage learners to produce language as they would when interacting with a human interlocutor during EFL class, learner input that was erroneous beyond a given threshold was detected and ‘rejected’ by the system. Thus, learner input with an error-to-word equal to or exceeding 25%, the dialog system issued a message asking the user to rephrase their statement. Only spelling errors were considered using the *LanguageTool* spellchecker (LanguageTool GmbH, 2013/2019), since spelling errors were the most frequent error types, as found in a previous pilot test of the dialog system, and because the tools detecting other error types were found not to be reliable enough to be implemented. As a note, the errors were not highlighted to the learner during data collection as this would create unequal conditions between learner-system interaction and learner-peer SCMC interaction, which had no error detection.

The tracker keeps track of the conversation state by storing user messages, executed system action (e.g., sending a message, listening for a user message), and slot values. Slots are predefined parameters that are filled with information the user provides. Operating as the system’s ‘memory’, the values of the slot are used by the dialog manager to decide on the next action to take. When the named entity recognition of the NLU extracts an entity from the user’s message, it is automatically stored in the corresponding slot. Entities that are extracted from the user’s messages are automatically stored in the corresponding slot.

The policy ensemble is a stack of three separate policies deciding on what the next action of the system should be. The first policy (i.e., Keras policy) predicts the action considering the intent and entities of the NLU, the previous system action and the current conversation state (i.e., slots). The second policy (i.e., Memoization policy) memorizes the conversations of the training data and overwrites the Keras policy prediction if it finds the same input in the training data. If both the Keras and the Memoization policy predictions have a confidence score below 40%, the Fallback

policy will issue a fallback message prompting the user to rephrase the previous statement.

If the prediction was above the configured threshold, the Natural Language Generator selected a response from a list of templates and the corresponding message was sent to the user. The response templates were created based on sample texts from the participants' English coursebook. Modeling the language of the system responses after the participants' textbook ensured that the language used was suitable to the learners' current language level and provided learners with model language input in line with the coursebook's and curriculum's target language norms for the text genre in question.

The responses were formulated based on an analysis of all sample texts of the genre service encounter (Koester, 2010) found in the learners' coursebook. Since the coursebook lacked sample texts for the communicative context chat, let alone chatbot interaction, the samples for e-mail and letter writing were included in the analysis instead, as they were considered more representative of the type of text suitable for a written chat interaction in the professional context than the samples of oral interaction (e.g., phone calls).

In the resulting corpus of six sample texts with a total of 535 words, all communicative functions matching one of the system's actions (i.e., confirm, offer additional service, request, thank) were extracted from the texts and compiled in a list (see Appendix B).

This overview provided a first reference point of how the communicative functions were realized linguistically in the learners' coursebook, e.g., high use of indirect formulations ("I can confirm that ...") and modal verbs ("could you let me know", "would be happy to"). Also notable are the absence of contracted forms such as "I'm" and exclamations.

A word frequency and concordance analysis using *AntConc* corpus analysis software (Anthony, 2019) revealed the specific linguistic patterns used in the corpus. Thus, high frequency of *would* and *could* (total 13 occurrences) confirmed the overall frequent use of modals. Furthermore, frequent use of *to* (28 occurrences) along with the use of infinitive *be* (eight occurrences) and the auxiliary *will* (four occurrences) indicated a high use of infinitive and passive constructions. With respect to the use of

pronouns, frequent occurrences of the second person pronouns *you* (21 occurrences) and *your* (10 occurrences) revealed frequent addressing of the recipient, higher occurrence of *we* (16 occurrences) and *our* (12 occurrences) compared to *I* (five occurrences) and *my* (0 occurrences) points to a preference of speaking in the first person plural over singular. There was also frequent use of *of* (13) in propositional phrases used as complements to nouns (e.g. “a minimum of three employees”) which along with the relatively low use of *that* (four occurrences) as subordinating conjunction to introduce a subordinate clause as in the example “I inform you that...” indicated a preference for the use of noun phrases over verb constructions (e.g., “arrange a viewing of the premise” rather than “arrange to view the premise”).

Based on these insights on the lexico-grammatical and stylistic features found in the coursebook sample texts, the system responses were created.

3.4.4 Training and Testing

The NLU and dialog manager require a sufficiently large set of training data to train their classification models. The NLU model is trained on example messages for each intent including entities. The training data should represent the range of possible user input for each intent so that the dialog system will later be able to appropriately map the many different linguistic variations to the corresponding intent.

To generate the training data for the dialog system developed here, the chat scripts from the testing of the task in pilot 1 were used as an initial corpus. This corpus consisted of seven interactions with a total of 273 messages. The messages sent by the student in the customer role were analyzed into 29 different intents (see Figure 4) and nine entities (see Figure 5).

Figure 4

Excerpt from List of Intents Derived from the Hotel Task Tested in Pilot 1

Intent	Examples
affirm	Yes Yes, please That is correct
ask_for_options	What rooms do you have? Do you have meeting rooms?
ask_for_room	Tell me about room Alpha What about the others? What about Alpha?
ask_for_room_equipment	Is it possible to show a presentation in Gamma?
greet	Hi there! Hey, how are you? Hello
greet+ask_for_options	Hello. What rooms do you have?
reserve_room	I want to reserve room Beta I want to book this room

Note. From “A Reduced List of Intents Derived From the Room Reservation Exercise of the Pilot Run” by Louis and Cocquio, 2019, *Chatbot in the English Classroom. Encourage Negotiations of Meaning* [Bachelor’s thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland], p. 12. FHNW. https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/thesis-cocquio_louis.pdf

Figure 5

Excerpt from List of Entities Derived from the Hotel Task Tested in Pilot 1

Entity	Examples
budget	1’500.- 1200 CHF 1500 1600 \$1200
date	25.04.2019 24 th of May 1 st December 2019
name	Yves Max Muster Mr. Muster
nr_of_people	150
room	Alpha Beta Gamma

Note. From “A Reduced List of Entities Derived From the Room Reservation Exercise of the Pilot Run” by Louis and Cocquio, 2019, *Chatbot in the English Classroom. Encourage Negotiations of Meaning* [Bachelor’s thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland], p. 12. FHNW. https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/thesis-cocquio_louis.pdf

Since the number of intent examples derived from the pilot data was not sufficient, additional examples were generated by the developers and researchers for both the training of the NLU as well as the dialog manager. To train the NLU, a domain-specific language (DSL) was used to generate various alternative formulations for each intent (see Figure 6).

Figure 6

Excerpt from the DSL File Used to Generate Alternative Formulations of Intents

```

%[ask_for_room_price]
  ~[whats] the price of ~[specific_room]?
  ~[is_there] ~[a_room] ~[for_singular] @[budget]?
  ~[can_we] ~[rent] ~[a_room] ~[for_singular] @[budget]?
  ~[is_there] ~[a_cheap] ~[room]?

~[can_we]
  Can we
  Could we
  Is it possible to
  Do you have

- What's the price of room [alpha](room)?
- Is there an even room for [1500 CHF](budget)?
- Is there a room for [1200.-](budget)?
- Can we book a room for [1500](budget)?
- Is it possible to reserve a room for [1500](budget)?
- Is there an inexpensive room?
- Do you have a cheap event room?
...

```

DSL file generates wordings

Note. From “An Excerpt of the DSL File Generating Wordings for the Intent Ask_For_Room_price” by Louis and Cocquio, 2019, *Chatbot in the English Classroom. Encourage Negotiations of Meaning* [Bachelor’s thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland], p. 25. FHNW. https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/thesis-cocquio_louis.pdf

To train the dialog manager, sample conversations between the user and the dialog system were generated (see Figure 7). Each conversation sample consisted of a sequence of user intents (including entities) and possible system responses.

Figure 7

Example of a Sample Conversation Used for the Training of the Dialog Manager

```
* greet
  - utter_greet
  - utter_ask_for_name
* provide_name{"name": "Yin"}
  - slot{"name": "Yin"}
  - utter_ask_for_service_with_name
* ask_for_room_size{"nr_of_people": "150"}
  - slot{"nr_of_people": "150"}
  - action_set_topic
  - slot{"topic": "size"}
  - utter_available_rooms_150_people
* ask_for_room_price
  - action_set_topic
  - slot{"topic": "price"}
  - utter_ask_for_room
```

Note. * intents. {} entities. - actions.

From “An Example of a Story Showing Intents (*) With Mentioned Entities ({key: value}) and Actions (-) the Bot Takes” by Louis and Cocquio (2019). *Chatbot in the English Classroom. Encourage Negotiations of Meaning* [Bachelor’s thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland]. FHNW.

https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/thesis-cocquio_louis.pdf

3.4.5 Usability Tests

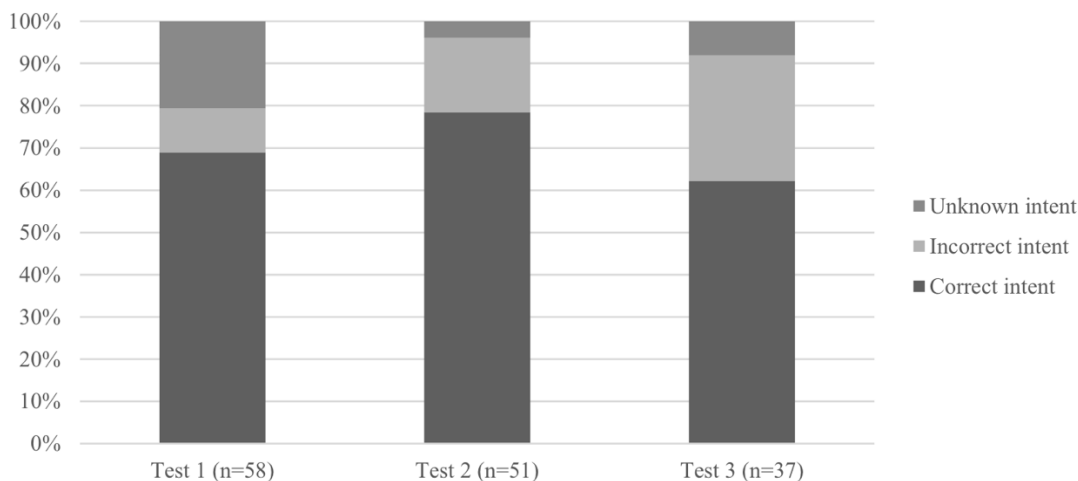
The developed dialog system was tested in three separate usability tests with three learners of the target group (i.e., students of two separate commercial VET schools) participating in each test. The participants were presented with the task and interacted with the dialog system on a test laptop. The developers were present in the room and simultaneously observed the system’s performance on a separate computer.

Across the three usability tests, 70% of the user messages (n = 146) were classified correctly (see Figure 8). Incorrect classifications occurred when the NLU was confronted with wordings that were not included in the training data or because different intents are similarly worded. In their evaluation of the usability tests, the system developers concluded that the NLU’s performance could be improved by training its model on a larger and more diverse training set collected directly from the target group rather than generating it (Louis & Cocquio, 2019).

Another source of error were intents that were unknown to the model, either because they were out of scope (e.g., “Which one of the rooms has been the most popular up until now?”) or when a message included several intents. In order to reduce errors due to unknown intents, the additional intent *ask_for_room_catering* and compound intents (i.e., intents consisting of more than one intent) for utterances containing the *greet* intent and an additional intent such as *ask_for_options* were introduced to the model after the first usability test. This reduced the incorrect classifications due to unknown intents in tests 2 and 3.

Figure 8

Classification of Intents by Dialog System in Usability Tests 1 – 3



Note. From “Proportions of Classified Intents per Usability Test” by Louis and Cocquio, 2019, *Chatbot in the English Classroom. Encourage Negotiations of Meaning* [Bachelor’s thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland], p. 27. FHNW.

https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/thesis-cocquio_louis.pdf

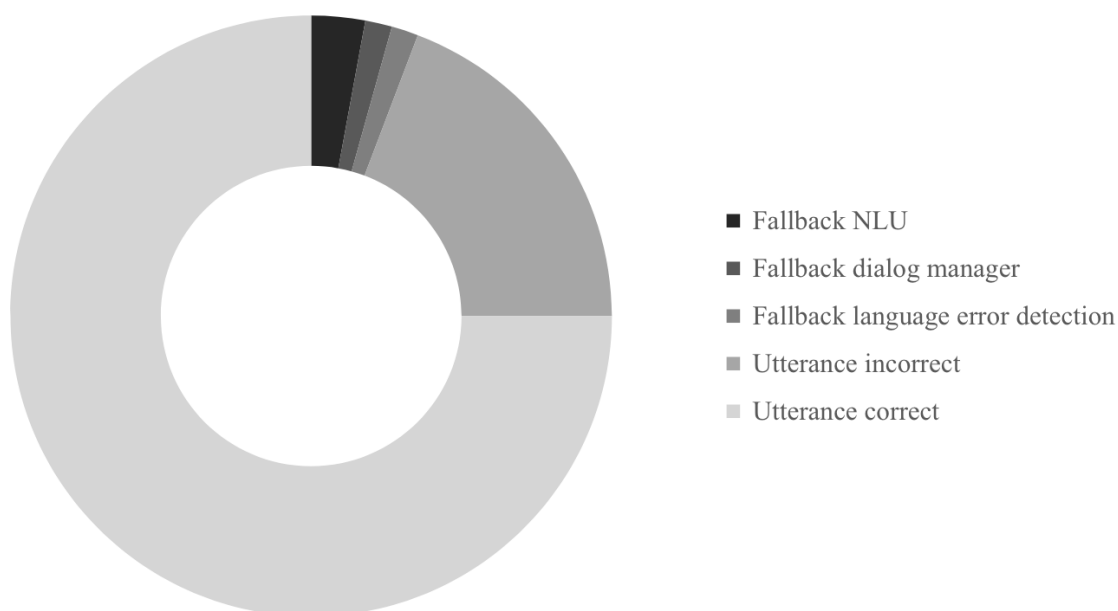
After processing the input from the NLU, the dialog manager issued a correct, i.e., appropriate, response in 75% of all system utterances (n = 208) (see Figure 9). The slightly higher success rate in correct system responses stems from the fact that the dialog manager decreased the impact of incorrectly classified intents. If the NLU for

instance confused *provide_nr_of_people* for *ask_for_room_size*, the dialog system would nonetheless provide an answer that makes sense to the user.

Nevertheless, across the three usability tests, 19% of the system utterances were incorrect, i.e., incoherent and nonsensical in the given context. In 5.6% of the system utterances, the fallback message was issued, triggered either by the NLU confidence threshold, the dialog manager, or language error detection.

Figure 9

Distribution of Dialog System Utterances Classified by Correctness and Type of Fallback Message



Note. From “Chatbot Utterances Labeled by Correctness and Origin of Fallback Messages” by Louis and Cocquio, 2019, *Chatbot in the English Classroom. Encourage Negotiations of Meaning* [Bachelor’s thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland], p. 28. FHNW.

https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/thesis-cocquio_louis.pdf

3.5 Pilot Study 2: Infrastructure, Questionnaires, and Interview

Pilot study 2 was conducted to verify the infrastructure (i.e., school internet network, laptops and other hardware, chat application, screen-recording software) and

pilot the additional instruments for data collection (i.e., questionnaires, stimulated recall) to be used in the main study.

Each participant was equipped with a test laptop to carry out the interaction task and complete the questionnaires. The laptops were connected wirelessly to the internet, and each had an external mouse. Each laptop was assigned to a participant so that the correct application (dialog system or chat application) could already be opened in the browser beforehand by the researcher: participants assigned to the system group interacted with the dialog system, which was a web application running in the internet browser (see section Dialog System), participants of the SCMC group interacted with their assigned peer interlocutor through the web chat application *Freenode* (Freenode, n.d.). Chatting channels for each student dyad of the SCMC group were also set up prior to the session. *OCam* screen-recording software (*OCam*, n.d.) was installed and running on all laptops to record the participants' screens during interaction. The participants were made aware that their screen activity would be recorded.

In addition to the handout stating the hotel task, the materials in pilot study 2 included a pre- and post-questionnaire that were being tested for the main study. Both questionnaires were designed by the researcher in an electronic web-based format using the survey tool *Unipark* (*Unipark*, 2019). The questionnaires were formulated in German.

The pre-questionnaire administered before the task contained two questions collecting the respondents' age and gender, six closed questions with frequency scale response format on the participants' previous chat and chatbot experience (e.g., "How often do you chat in English?", "How often do you chat or speak with robots and other devices?"), and a final open-response question for general comments.

The post-questionnaire was administered immediately after the interactions took place and collected the students' perspectives on the task. It included 17 close-ended items with scaled, multiple-choice, and yes/no-response formats, with three of them being followed by an additional question with open-response, as well as a final open-response question for general comments. Some item examples are: "I enjoyed the task", "The task is useful for English learning", "Did you understand your conversation partner?", "What did you do when you couldn't understand your conversation

partner?”, “In what style did you write?”, and “What style of writing would be appropriate for the task?”. The final versions of both questionnaires can be found in the appendix (see Appendix D).

The interview guide was developed for use during interviews with select participants in session 3, one week after the interactions. The guides were designed following the recommendations on stimulated recall by Gass and Mackey (2016), a method commonly used in L2 and FL research to “uncover the cognitive processes which might not be evident through simple observations” (Gass & Mackey, 2016, p. 25). In this study, the method was used to access the participants’ thought processes during their interaction, with the goal to be able to describe the aspects of interaction found in the chat scripts more comprehensibly.

The first part of the guide consisted of a word-by-word script for the researcher’s oral instructions to the participants on the purpose and procedure of the interview. When formulating the script, attention was paid to highlighting that participants should recall the events in the previous week, rather than comment on their reflections at the time of the interview. The script further explicitly stated that the researcher did not favor any particular kind of response, ensuring that the students’ actual thought processes and strategies were tapped into (Gass & Mackey, 2016).

The main part of the interview guide consisted of the stimuli for the recall activity. This part of the guide was finalized after the interactions had taken place and the stimuli had been selected for each interviewed participant from the chat scripts and screen-recording videos. The selection process will be explained below. In its final form, the recall part consisted of five to twelve stimuli, along with suitable question prompts. As Gass and Mackey (2016) point out, the questions asked during a stimulated-recall interview can “potentially influence and compromise the procedure” (p. 50), for instance by implying the answer or eliciting a response that the participant believes the interviewer expects or prefers. To prevent such influence, the questions were standardized for all stimuli. Furthermore, the questions were phrased in past tense so as to elicit the thoughts the participant had during the event, preventing “intervening thoughts” involving reflection at the time of the interview about the event (Gass & Mackey, 2016, p. 50).

The final part of the interview guide consisted of a semi-structured interview guide with five open-ended questions, including back-up prompts on the students' perception of the task and their interaction, more precisely, their general perception of the task, their own and their interlocutor's communication, and their thought processes with regard to task completion.

An example of the final interview guide used in the main study can be found in the appendix (see Appendix E).

Pilot 2 took place on two separate days one week apart in a class of eleven third-semester students from the school of the main study. Prior to the data collection, all participants were anonymized with a six-digit code which could be used simultaneously as authentication ID for the dialog system application. The code served to later match the same participant's data from the different data sources. The students were randomly assigned to a learner-peer SCMC (6 participants) or learner-system (5 participants) group. Within the learner-peer SCMC group, the participants were further randomly assigned into dyads.

The first session was conducted in a classroom equipped with test laptops. The data collection was led by the researcher, with the teacher being present in the room. The researcher informed the participants about the study procedure following a script. Participants subsequently gave their written consent to participating in the study.

The class then simultaneously read through each item in the pre-questionnaire and gave oral feedback in their native dialect, Swiss German, on the intended comprehensibility and relevance of each item. The researcher took notes and audio-recorded the conversation with the participants' consent.

After completing the pre-questionnaire, the students were handed out the task and given the opportunity to ask questions. They then carried out the task: participants of the system group interacted in the role of hotel customer with the dialog system simulating a hotel chatbot, participants of the SCMC group interacted in the role of hotel customer with their assigned peer interlocutor who played the role of hotel employee.

When starting chatting, two students experienced trouble using the chat application and were instructed by the researcher to carry out the task with the dialog system instead. The researcher assisted with technical issues when necessary. Time

was limited to maximum of 20 minutes and students were instructed to complete a grammar worksheet in case they completed the task before the time was up. After 20 minutes, the participants completed the post-questionnaire, following the same procedure as in the pre-questionnaire.

After this first session of the data collection, the researcher exported the four SCMC chat scripts and seven learner-system interaction scripts along with the screen-recording videos from the participants' laptops. The data were imported into MAXQDA 2020 (VERBI Software, 2020) and roughly coded for sequences pertinent to the research question with the goal to extract them as stimuli for the recall part of the retrospective interviews.

Since the available time for conducting the interviews was limited to 90 minutes due to the school schedule, a total of four participants could be selected for interviewing. One participant from the SCMC group and three participants from the system group were selected based on the density and uniqueness of relevant sequences in their interactions, i.e., sequences relating to the research questions as described above. For each participant, the relevant segments were extracted from the chat scripts and screen-recording videos. A screenshot of each segment was placed in the interview protocol along with a suitable prompting question, as outlined above. This procedure resulted in a personalized interview guide for each participant with five to twelve segments and corresponding question prompts.

The interviews were conducted at the earliest possibility that the school schedule allowed, which was one week after the interactions took place. It is generally recommended to collect stimulated recall data "as soon as possible" (Gass & Mackey, 2016, p. 52), and ideally, less than 48 hours after the task itself in order to ensure best possible recollection of the event. However, measures such as choosing a "strong" stimulus (Gass & Mackey, 2016, p. 44) that is close to the event itself and overall careful adherence to the guidelines for stimulated recall stated above, should help minimizing possible effects the timing may have on the participants' memory. Possible memory decay must also be taken into account as potential limitations when interpreting the participants' answers.

The interviews were conducted in Standard German. The time was standardized to about 20 minutes per interview due to the time restrictions mentioned above. At the

beginning of the interview, the researcher instructed the participant on the interview procedure following the script.

During the stimulated-recall part, the researcher presented the selected stimuli to the participants one at a time by playing the video segment on a laptop or showing a highlighted segment on the script printout, respectively. When prompting the recall, careful attention was paid to adhering to the standardized question format, avoiding any back-channeling, and allowing the participants enough time to complete the recall. Whenever the students' response was unclear with respect to the time of reference, the researcher asked the participant to specify whether they were in fact commenting on their thoughts they had the week before or instead, their interpretation at the time of the interview. After the response to each question, the interviewer waited for a few seconds before asking the next question. At completion of each segment recall, participants were asked if there was anything else that they remembered.

The researcher took notes and audio-recorded the conversations with the participants' prior consent, paying careful attention to reacting to all responses in a similar manner, thus preventing the participants from assuming that any aspects of their responses are more relevant than others.

The dialog system proved to be able to manage the simultaneous use by multiple users. However, the chat application *Freenode* (Freenode, n.d.) was found not to be suitable. For school infrastructural reasons, the chat application only allowed a limited number of users to simultaneously use the application. Therefore, an alternative application had to be selected for the main study, which will be described below.

The task was carried out as intended. However, the interaction scripts revealed that some students showed a tendency to use an informal rather than a more formal style as would be appropriate for the professional context. To encourage learners to consider more consciously the style suitable to the communicative context of the task, a prompt was added to the task handout asking students to write in a manner that is suitable for the professional context.

Both questionnaires were adapted according to the students' oral feedback. Adaptations included reordering and reformulations of questions and changes to the response formats (e.g., multiple-choice rather than open-ended format).

The interview guide and procedure were slightly adapted for implementation in the main data collection as follows. Rather than two different types of prompts (i.e., script printouts and video segments), participants would be presented with video segments only during the main data collection. It was assumed that switching between two different types of stimuli would potentially guide the participants' attention towards a particular aspect or answer pertinent to the medium. Video rather than the printed script was chosen for the main study because it is considered the stronger stimulus as it is more similar to the actual event. Moreover, it became evident during the pilot interviews that some students appeared to be hindered from verbalizing their thoughts in Standard German as opposed to their native dialect Swiss German. It was therefore decided that participants would be given the option to hold the interview in Swiss German during the main data collection. The semi-structured interview questions were deemed suitable and conducive to eliciting the participants' thoughts on the issues relating to the research question. With the goal of collecting any additional own observations and insights, a final question was added, asking participants whether there was anything else that had not been addressed during the conversation.

3.6 Pilot Study 3: Final Questionnaires and Tasks

A final pilot study was conducted to finalize the student questionnaires and pilot the final version of the task. The evaluated materials contained the pre- and post-questionnaires and task (dialog system group only) which included the amendments from pilot study 2.

The pilot study was conducted at a commercial VET school different from the one where the main study and other pilot studies took place in one first-year class (n = 17) and one third-year class (n = 16). Each student used a desktop computer with wired internet connection in a school IT room. The procedure was otherwise identical to pilot study 2.

After data collection in one class, the researcher amended the questionnaires according to the students' feedback. During data collection in the second class one week later, the participants gave feedback on the amended version of the questionnaires.

Both questionnaires were further adapted and finalized for the main study according to the students' feedback. Adaptions included reformulations and additions of examples.

3.7 Main Study

3.7.1 Participants

The participants consisted of a total of 39 students (21 female, 18 male) from three first-year classes at a commercial VET school in the Swiss canton of Bern. The participants were recruited through their English teachers with permission of the school management. The study took place during English class and participation was voluntary for the students. Prior to the study, all participants signed an informed consent (see Appendix G).

Participants' ages ranged from 16 to 27 years ($M = 17.41$; $SD = 1.85$). As in pilot study 2, each participant was anonymized using a code that could simultaneously be used as authentication for the dialog system application.

At the time of the study, the students were midway through their second semester of the three-year program. During their commercial VET program thus far, they had been following English classes for three lessons per week and had typically previously attended five years of English as a foreign language during compulsory schooling. The three participating classes were taught by two different English teachers who followed the same curriculum and used the same coursebooks.

The participants' English proficiency levels ranged from A2 to C1 as assessed for the purpose of this study with UCLES 2001 Quick Placement Test paper and pen version (University of Cambridge. Local Examinations Syndicate, 2001). Table 2 presents the number of participants that reached each of the four levels. As can be seen, the majority were at levels B1 or B2, with about half reaching level B1 and a third

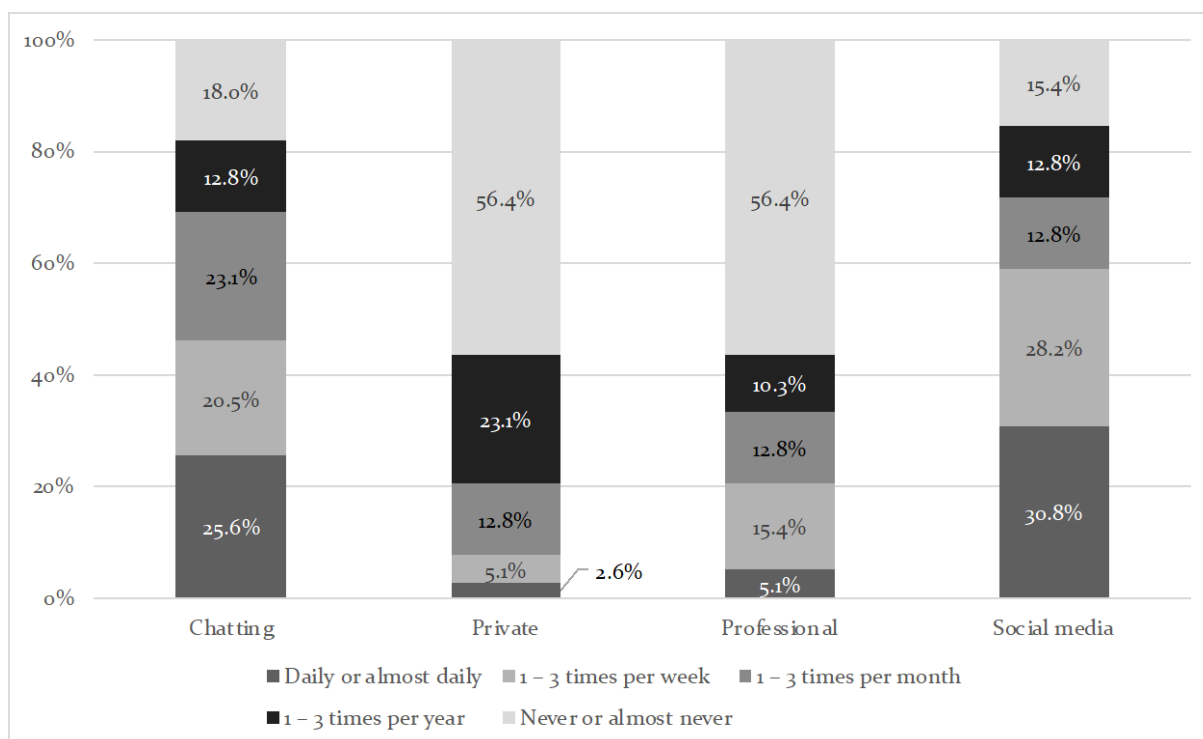
reaching level B2. A few learners reached level A2. Level C1 was reached by one participant.

Table 2

Participants' Language Levels

Level	<i>n</i>	%
A2	5	12.8
B1	20	51.3
B2	13	33.3
C1	1	2.6

In terms of previous experience writing and chatting in English, the group was heterogeneous. As shown in Figure 10, about half of the participants indicated that they chatted in English frequently, i.e., daily or once to three times per week. The other half used English chat rarely, i.e., once to three times per month or once to three times per year, or never. In terms of other forms of written communication, most participants never or very rarely wrote emails or letters in English: For private written communication, only a few participants used English daily or once to three times weekly, while the majority seldom wrote in English for private purposes, i.e., once to three times per month, once to three times per year, or never/almost never. In the professional context, a minority of participants used English regularly, i.e., daily or almost daily or once to three times per week. The majority rarely wrote in English, i.e., never/almost never, once to three times per year, or once to three times per month. However, the use of written English was higher in the context of social media such as posting and commenting in social media networks, blogs, or forums. About half of the participants used English in social media daily or once to three times a week. The other half did so once to three times per month, once to three times per year, or never.

Figure 10*Participants' Frequency of Writing and Chatting in English*

Note. $N = 39$

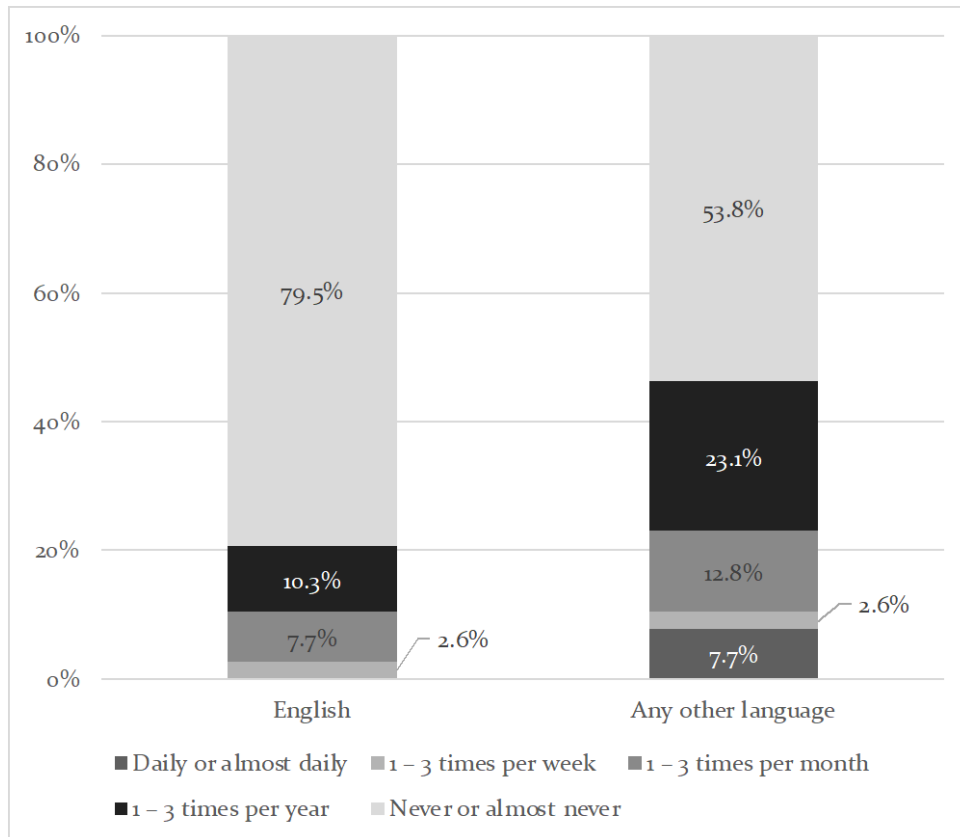
Only a few participants reported that they regularly communicated with dialog systems. Communication with dialog system was specified as written and spoken interaction with virtual assistants built into smartphones (e.g., *Siri*, *Google Assistant*) or other devices such as speakers (e.g., *Amazon Alexa/Echo*, *Google Home*), chatbots, and robots. The participants' responses about their dialog system use are summarized in Figure 11. Without specifying the language used, a few participants reported daily or weekly interaction with dialog systems. About half of the participants never communicated with dialog systems and the remaining participants did so infrequently, i.e., once to three times per year or once to three times per month.

Even fewer participants reported regular experience communicating with dialog systems in English. No participants indicated daily use, while one participant reported to communicate with dialog systems in English once to three times per week. About a third of the participants communicated with dialog systems in English rarely, i.e., once

to three times per month or once to three times per year. Most participants reported that they never communicated with dialog systems in English.

Figure 11

Participants' Frequency of Communicating with Dialog Systems in English or Any Other Language



Note. $N = 39$

With respect to the participants' previous experience using computers and keyboards, it could be assumed that they all had regular experience, given that the participants were in training as commercial employees.

3.7.2 Materials

The materials included instructor's scripts for each data collection session (see Appendix F), student consent forms (see Appendix G), the paper and pen version of

the *UCLES 2001 Quick Placement Test* (University of Cambridge. Local Examinations Syndicate, 2001), and the finalized versions of the task handouts (see Appendix C), pre- and post-questionnaires (see Appendix D), and interview guides (see Appendix E).

3.7.3 Infrastructure

The infrastructure was the same as in pilot 2 described above: Each participant was equipped with a test laptop with wireless internet connection and an external mouse. As in pilot 2, the laptops had been previously assigned to the participants with the corresponding application (dialog system or chat application) already open in the browser and chatting channels for each student dyad of the SCMC group set up. For the SCMC group the open-source chat application *Mattermost* (*Mattermost*, 2019) was used. The browser's spellchecker was deactivated on all laptops.

As in pilot 2, *Ocam* (*Ocam*, n.d.) screen-recording software was installed and running on all laptops with the participants' consent.

Figure 12

User Interface of the Dialog System Application

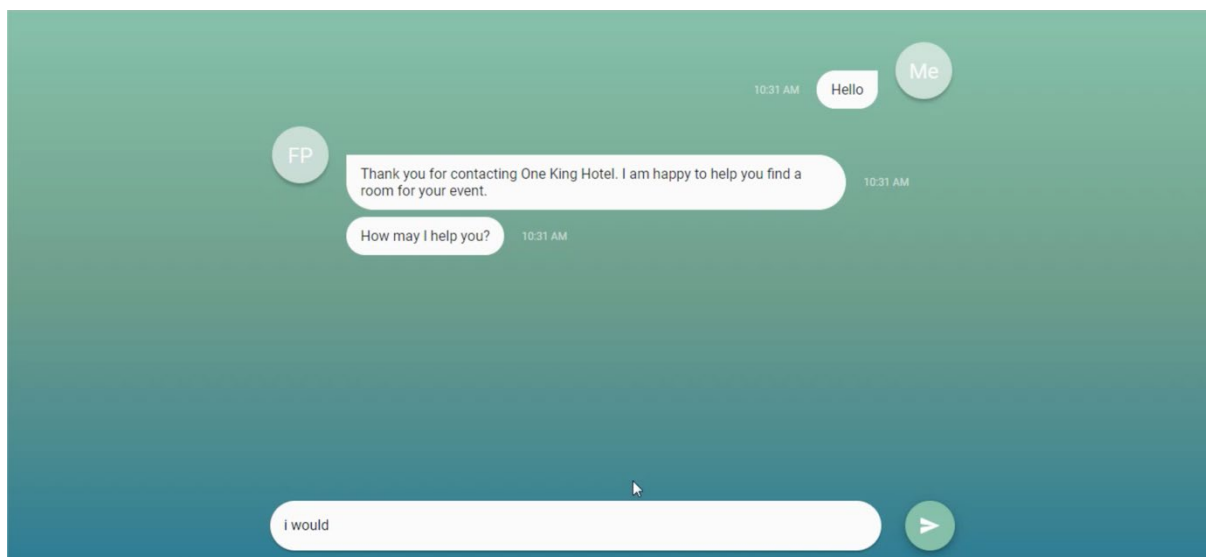
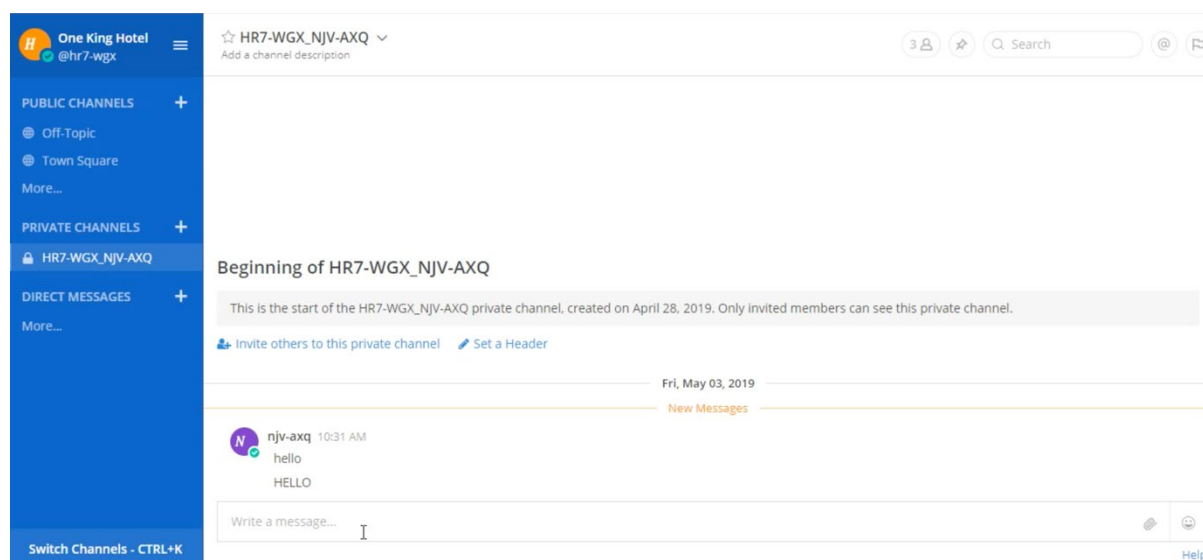


Figure 13*User Interface of the SCMC Application***3.7.4 Procedure**

Data collection was conducted in three sessions on separate days with one week between the sessions.

Session 1: Study Introduction and Language Test

The first session took place in the participants' regular English classrooms. The researcher informed the participants on the purpose and procedure of the study, following the script. All students were present during Session 1, and all signed the consent form. The participants then completed the language test. Time was limited to the test's intended duration of 30 minutes. Students who finished the test early were handed part 2 of the test. When determining the proficiency level, only part 1 was considered. The language tests were marked by the author in preparation of Session 2.

Session 2: Interaction Tasks and Questionnaires

Session 2 was held one week later in the school auditorium with the teacher being present in the room. The room was equipped with rows of tables where the laptops were placed.

Following the same procedure as in pilot study 2, the students of each class were randomly assigned to the SCMC (a total of 18 students across all three classes) or learner-system (a total of 21 students across all three classes) chat group. The participants of the SCMC chat group were assigned into dyads based on their language proficiency, matching students with similar levels of English to each other. This was done to prevent more advanced students from getting frustrated with the slower response time by their less fluent interlocutor or learners from feeling intimidated by their more fluent chat partner. Since one student was absent for session 2 without prior notice, the participants had to be reassigned to the two groups and, if part of the SCMC group, another chat partner immediately before the start of the session.

The researcher informed the participants about the procedure following the instructor's script. The participants then completed the pre-questionnaire. When all participants had completed the questionnaire, they moved on to reading the task handout. The students had the opportunity to ask questions about the task and subsequently carried out the task. The procedure during the task interactions was identical to pilot study 2. After the time for the task was up, students completed the post-questionnaire.

After the session, the researcher followed the same procedure as in pilot study 2, transferring the scripts and screen-recording videos from the student laptops. Any student names appearing in the scripts were replaced with pseudonyms.

As described above, only video segments were to be used as stimuli for the stimulated-recall interviews in the main data collection. Consequently, only the videos were coded using the same codes as in pilot study 2, adding *alignment* as a new code that emerged from the data. After the coding, the most pertinent interactions and video sequences were selected for the stimulated-recall interviews and used to prepare the interview protocols.

Session 3: Interviews

The interviews were conducted one week after the chat interactions took place. Due to the school schedule, interviewing time was set to 90 minutes per class or 20 minutes for each participant. From each class, an equal number of participants from

the SCMC and learner-system chat group (six SCMC, six learner-system) were selected based on the considerations as described for pilot study 2.

Beginning the interview, the researcher stated the instructions in Standard German. The participants were then given the option to continue the interview in Swiss German. As mentioned above, it was assumed that learners would feel more at ease verbalizing their thoughts in their native dialect Swiss German rather than Standard German. Except for one student, all participants chose to continue the interview in Swiss German. Conducting the interviews, the researcher adhered to the same guidelines as specified above with regard to pilot study 2. The interviews were audio-recorded with the participants' prior consent.

3.7.5 Data Processing and Analysis

Coding of Interactional Features in Scripts

To address research question 1, the interaction scripts were coded with the aim of identifying interactional features and their specific characteristics as they occurred in learner-system interaction. The following sections describe the procedure that was followed to this aim. An introductory paragraph explains the general coding procedure. This is followed by a detailed description of the definitions and guidelines for the coding of each interactional feature. The procedure followed to ensure intercoder reliability is documented at the end of the chapter.

General Coding Procedure

Since the focus of the analysis was on interactional features and their characteristics as they occur in the learner-system interaction type, the codes relating to interactional features were established based on the learner-system interactions and subsequently adapted to accommodate the differences of the SCMC interactions. The adaptations made are documented below.

An initial code system for interactional features was developed by the author based on a separate sample of eight learner-system interactions from a pilot study. The development of the code system for interactional features followed a top-

down/deductive and bottom-up/inductive approach: As a general procedure, interactional features were first identified and classified into main codes following the definitions used in the existing literature. This yielded three types of interactional features in the learner-system interactions: negotiation of meaning, recasts, and lexical and structural alignment. The main codes were then refined into subcodes by applying existing codes from the literature as far as possible or developing new codes for phenomena not previously categorized in the literature, but which were considered relevant with regard to the efficacy of negotiation of meaning to promote EFL learning. The following paragraphs describe the coding procedure relating to each interactional feature in detail, also highlighting which codes were adopted or adapted from the existing literature and where the need emerged to develop new codes.

The codes themselves, both those taken from the literature as well as those newly developed, will be presented in detail and illustrated with examples from the data in the chapter Findings. This organization is chosen since the newly developed codes pertain to phenomena specific to the characteristics of learner-system interaction and are thus considered a finding of the study. Similarly, applying the existing codes, which had been established based on human-human interaction, to the data revealed that while they do fit the interactional features in learner-system interaction, their specific linguistic realization differs, sometimes considerably, from that found in the literature on human-human interaction. Since these differences are considered a finding relevant to the research question, they will be illustrated with examples in the chapter Findings as well.

As mentioned above, the code system established based on the learner-system interactions was subsequently applied to a sample of five SCMC scripts from a pilot study and slightly adapted to account for the differences of the interaction modality. The adaptations to the original learner-system code system when applied to the SCMC scripts involved the following two changes: First, the additional negotiation of meaning subcode *overlap* was added. The nature and occurrence of the new code will be described in the chapter Findings. As a note on terminology, *learner* is consistently used to refer to the student in the role of the customer, whereas *interlocutor* refers to the dialog system in learner-system interaction and the student peer in SCMC who were in the role of the hotel employee.

Second, the coding procedure for the main codes *requests* and *recasts* had to be slightly adjusted to fit the SCMC context. More precisely, the coding guidelines of *requests* and *recasts* were complemented with a specification that accounted for the different message structuring behavior of learners when chatting with a peer instead of the dialog system: In the SCMC group, learners frequently split utterances over multiple messages. As a consequence, recasts and requests in the SCMC interactions were sometimes split over several messages. However, in the system group, learners were prevented from splitting their utterances into multiple messages as the dialog system by design responded to each message immediately after it is sent by the learner. In other words, a learner's turn in the learner-system interaction automatically consists of a single learner message. The code guidelines for *requests* and *recast* were adapted accordingly, allowing for application of the code across multiple messages.

As a final adaptation of the code system, guidelines for the coding of *lexical* and *structural alignment* were adapted to the SCMC context: Since the SCMC chat interface typically showed ten rather than five messages at any given time, the threshold for the coding of alignment in the SCMC interactions was extended from five (learner-system context) to ten messages (SCMC context).

The resulting code system was subsequently applied to the two sets of interaction scripts (system, SCMC). For ease of coding, the code system was applied in several steps, as the coding of some codes relied on the pre-coding of other codes (e.g., recasts required pre-coding of errors), as will be described in detail below (see section Intercoder Reliability). Furthermore, coding was conducted by two raters to ensure intercoder reliability, which will also be detailed below (see section Intercoder Reliability).

Coding of Negotiation of Meaning

Episodes of negotiation of meaning were identified using the model by Varonis and Gass (1985), i.e., as sequences where “some overt indication” of non-understanding occurs between the interlocutors and the conversation is “pushed down” while the interlocutors attempt to resolve the communication breakdown (p. 73). According to the model, negotiation of meaning sequences consist of four components (see chapter

Literature Review for detailed description of model): trigger, indicator, responses, and the optional reaction to response.

Following previous studies on negotiation of meaning using the same model (e.g., Nakahama et al., 2001), negotiation episodes were identified by first locating the indicator. Both negotiation sequences initiated by the learner and by the dialog system were considered. As a next step, each negotiation component as well as the negotiation sequence as a whole were further refined into subcodes. As mentioned in the previous paragraph, subcodes were adopted or adapted from existing codes in the literature or they were newly created when no suitable codes for the observed patterns were described in the literature.

When developing new subcodes, the focus was on aspects that relate to concepts which are associated with EFL acquisition specified by cognitive-interactionist theory. So, according to the theoretical claims about negotiation of meaning sequences, the subcodes were defined with the goal to describe to what extent the negotiation sequence provides the learner with opportunities to receive comprehensible input and facilitate noticing and the production of modified output and repair.

Table 3 shows an overview of the main and subcodes along with their sources used for the coding of negotiation of meaning sequences. Subcodes with no source indication were developed anew. As stated in the previous paragraph, all subcodes will be presented in detail along with examples in the chapter Findings. Detailed code descriptions and examples can also be found in the codebook for interactional features (for link, see Appendix H).

Table 3*Overview of Main and Subcodes for Negotiation of Meaning*

Main code	Subcodes
Trigger	
Accuracy	Erroneous Target-like
Indicator	
Initiator	Learner Interlocutor
Type	Clarification request (Long, 1983) Confirmation check (Long, 1983) Inappropriate response (Pellettieri, 2000; Varonis & Gass, 1985) Overlap
Response	
Type	Meta-comment Change topic Accept topic (adapted from Long, 1983) Elaboration (Smith, 2003; Varonis & Gass, 1985) Reduction (Varonis & Gass, 1985) Modification Repetition (Long, 1983; Smith, 2003; Varonis & Gass, 1985) No response
Accuracy of reformulation	Improved Reduced Equivalent
Reaction to response	Implicit resolution Explicit resolution Implicit no resolution Explicit no resolution No reaction to response
Complexity	Single-layered (Chen, 2018) Multi-layered (Chen, 2018)

Note. References indicate the source of the respective subcode. Subcodes without indication of source were developed anew.

Coding of Recasts

Recast episodes were defined following Long (1996) and Sheen (2006) as “episodes in which the interlocutors rephrase a learner utterance to be more target-like by changing one or more sentence components while retaining its central meaning and without breaking the flow of the communication” (Smith, 2010, p. 88).

It needs to be noted that in the learner-system interaction type the interlocutor, i.e., the dialog system, self-evidently did not actively and willingly rephrase an

utterance as human native speakers or teachers do while attempting to provide feedback to a language learner. Rather, the dialog system's utterance following a learner's utterance might happen to contain a more target-like formulation. However, since the concept was consciously adopted from the context of human-human interaction and for comparability reasons, the formulation of the definition was maintained.

As mentioned above, for ease of coding, recasts were coded after errors had been coded. The coding of errors will be detailed below (Coding of Errors). After pre-coding the errors, the main code *recast* was applied following the definition and guidelines in the codebook (for link, see Appendix H).

After the recast segments in the scripts were identified, their two components *recast move* and *uptake* (see chapter Literature Review) were further refined into subcodes. To categorize the recast moves into subcodes the taxonomy by Sheen (2006) served as basis. Some of the subcodes were adopted unchanged, some subcodes were adapted to fit the specific nature of the recasts found in the data, and some additional subcodes were newly developed. When developing new subcodes, the focus was on the aspects salience and explicitness, the two major contributors to the effectiveness of recasts in promoting FL learning: As outlined in the literature review, the efficacy of recasts depends on the extent to which they indicate to the learner that they have made an error, highlight what the error is, and make the target-like form salient. Similarly, the uptake component was categorized into subcodes related to the efficiency of the recast, hence its ability to make the learner aware of their error and enabling the use of the target-like form. The subcodes were adopted or adapted from the existing literature (Lyster & Ranta, 1997; Smith, 2005) or developed anew where necessary. Table 4 provides an overview of all main and subcodes used for the coding of recasts along with their sources. All subcodes will be presented in detail along with examples in the chapter Findings. Detailed code descriptions and examples can also be found in the codebook (for link, see Appendix H).

Table 4
Overview of Main and Subcodes for Recasts

Main code	Subcode
Recast move	
Linguistic focus	Spelling Lexis (Sheen, 2006) Grammar (Sheen, 2006) Multi-focus
Intensity of recast	Singular incidence Repeated incidence
Intensity of recast form	Singular incidence Repeated incidence
Completeness	Complete Partial
Length	Single sentence Multiple sentences Multiple messages
Uptake	No uptake (Lyster & Ranta, 1997) Immediate uptake (Smith, 2005) Repair (Lyster & Ranta, 1997) Needs repair (Lyster & Ranta, 1997) Delayed uptake (Smith, 2005) Repair Needs repair 'Re-uptake' Repair Needs repair

Note. References indicate the source of the respective subcode. Subcodes without indication of source were developed anew.

Coding of Alignment

Alignment was defined in line with previous FL research as the repetition of a linguistic feature (i.e., lexical material, grammatical form) used in previous discourse as the result of priming (Michel & Cappellini, 2019). Lexical and structural alignment were differentiated and coded separately following the operationalizations and procedures outlined below.

Coding of Lexical Alignment

Following the procedure outlined in Michel and Smith (2017) and Michel and Cappellini (2019), who explored alignment between FL learners in SCMC, instances of

lexical alignment were identified by initially identifying and coding lexical overlap of sequences of three or more consecutive words, i.e., n-grams, in the scripts. To this goal, the scripts were first spellchecked and spelling errors were manually corrected. The scripts were then imported into *AntConc* concordance software (Anthony, 2019). Only n-grams that the software found at least twice within the script and that showed exact overlap were considered for further analysis (Michel & Cappellini, 2019). Since the main interest of the analysis was on the learners' alignment behavior rather than the interlocutor's (i.e., dialog system or peer, respectively), repeated n-grams that appeared exclusively in the interlocutor's production were excluded. N-grams essentially consisting of room names such as, e.g., *the room Beta* were also excluded.

Once the n-grams were identified in the scripts as potential sources for lexical alignment, they were manually searched and coded in MAXQDA 2020 (VERBI Software, 2020) applying the main code *lexical alignment*. This preparatory work was done by the author.

The subsequent assigning of subcodes was conducted by both the author and the second coder for a sample of scripts and intercoder agreement was checked to ensure reliability, as described below (see section Intercoder Reliability). As a first coding step, it was decided for each n-gram pre-coded as *lexical alignment* whether it was considered alignment or not: An occurrence of an n-gram was considered alignment when the same n-gram occurred within the five preceding messages sent by either the learner or the dialog system.

Example 1 illustrates an instance of lexical alignment according to this operationalization. In the example, the 4-gram "tell me which room" from the dialog system's utterance in message 25 is reused with exact overlap by the learner three messages further down, in message 28.

Example 1*Coding of an n-gram as Lexical alignment*

- | | | | |
|----|---------|--|--------------------------|
| 25 | System | Could you please <u>tell me which room</u> you are referring to? | |
| 26 | Learner | i like all 3 rooms, can you help me to decide? | |
| 27 | System | So sorry, I could not understand your last message. Could you please say it again? | |
| 28 | Learner | can you <u>tell me, which room</u> is the best for my options? | <i>Lexical alignment</i> |

However, if the n-gram was reused outside the five-message threshold it was coded as *unprimed*. Example 2 illustrates the unprimed reuse of an n-gram. In this example, the learner reuses in message 12 their own expression “be able to play video and audio” from message 1. However, since the expression occurs outside the five-message threshold, it is considered unprimed reuse.

Example 2*Coding of an n-gram as Unprimed*

- | | | | |
|--------|---------|--|-----------------|
| 1 | Learner | Hello, i am searching for a room suitable for our company event that costs 1'500 Francs maximum. It will take place on 12th October 2019. The room should be big enough to fit a minimum of 150 people. We also need to <u>be able to play video and audio in the room</u> . | |
| 2 – 11 | | [...] | |
| 12 | Learner | Will we <u>be able to play video and audio in room alpha</u> ? | <i>Unprimed</i> |

The five-message threshold was established following Michel and Cappellini (2019). For their analysis of the alignment in learner-peer written SCMC, the authors specified an eight-message threshold for their data, which corresponded to the number of messages visible to the learner on the screen at any given time. The authors based their rationale on previous alignment analysis in FTF spoken interaction, specifically the study by Dao et al. (2018), where the threshold was set to eight turns. This in turn

was built on the idea from FTF spoken interaction research that the “activation of a prime would decay over turns after a certain time” (Michel & Cappellini, 2019, p. 196). Since the dialog system interface used in this study typically showed five messages to the learner, this threshold was adopted for the analysis.

However, the procedure followed in the present study where the coders searched ‘backwards’ starting from an item pre-coded as *lexical alignment* and searching lexical overlap in preceding utterances, differs from the one proposed by Dao et al. (2018) and Michel and Cappellini (2019). In their analyses, Dao et al. (2018) and Michel and Cappellini (2019) worked through the scripts chronologically, coding each first occurrence as *prime* and any subsequent occurrence as *target*, given that it was within the specified threshold. This procedure results in *prime-target* sequences with the first item being coded as *prime* and any subsequent occurrences within the threshold as *target*, preventing an occurrence from being considered simultaneously a prime and a target. The procedure had been adopted from priming experiments (specifically, H. P. Branigan et al., 2000) which typically examined dialogues that consisted of two-turn sequences with the first turn containing the prime and the second the target. For the present study, however, this was considered an unsuitable procedure since the scripts represented continuous interactions where alignment sequences may show multiple occurrences of a repeated n-gram and run over several turns, frequently extending beyond the five-message threshold. For the present data, it was therefore deemed inadequate to not consider a target as a potential prime and alignment was coded backwards, as described above.

The unprimed use of n-grams was further classified as done by the learner or by the interlocutor. However, the direction of the unprimed use was not considered since it would be inconsistent to consider the alignment unprimed and at the same time primed by one of the two types.

Following previous literature on lexical alignment (Dao et al., 2018; Michel & Cappellini, 2019), the instances identified as alignment, as described above, were assigned further subcodes specifying the directions of lexical alignment, i.e., from learner to learner, from learner to interlocutor etc. Table 5 shows an overview of the subcodes used to describe the different directions of lexical alignment along with examples.

Table 5
Overview of Subcodes for Lexical Alignment

Subcode	Example
Learner to learner (self)	Learner: <u>What's the price for room Gamma</u> System: Alpha costs CHF 1400. Learner: <u>What's the price for room beta?</u>
Learner to interlocutor	Learner: Thank you for your reservation and <u>have a nice day</u> . System: <u>have a nice day</u> .
Learner to both	Learner: That is perfect. I <u>would like to book room Alpha for the 12th of October 2019</u> then, if that's possible. System: Please could you confirm if you <u>would like to book room Alpha for the 12th of October 2019?</u> Learner: I hereby confirm, that I <u>would like to book room Alpha for the 12th of October 2019</u> .
Learner unprimed	Learner: okey which room is the best for <u>a good atmosphere</u> [Messages 15 – 22] [...] Learner: in which room can i have <u>a good atmosphere?</u>
Interlocutor to learner/With error	Learner: the date is on <u>the 12 oktober</u> System: I will reserve room Gamma for <u>the 12 oktober</u> . Could you please confirm if this is correct?
Interlocutor to learner/Error-free	Learner: What's the price <u>for room Gamma</u> System: The price is CHF 1100 CHF <u>for room Gamma</u> .
Interlocutor unprimed	System: Very well. Would you <u>like to book</u> room Gamma? [Messages 38 – 48] [...] Learner: I would <u>like to book</u> the Beta room

As can be seen in Table 5, instances where the interlocutor aligned to the learner were also coded. Considering the theoretical presumption that alignment is an inherently sociopsychological phenomenon, it might seem controversial to apply this concept to a dialog system's discourse 'behavior'. While this study does not assume that the same processes are at work when the dialog system 'aligns' to the learner, it can be hypothesized that the effect on learning is the same as when done by a human interlocutor.

Furthermore, as also seen in Table 5, instances that were assigned the code *interlocutor to learner* were further classified as *with error* or *without error*. This step was done to be able to highlight instances where the interlocutor repeated a learner error, providing the learner with inadequate feedback and potentially reversing the positive effects of interlocutor alignment for the learner.

Instances where the interlocutor reused lexical material from their own production, i.e., alignment from interlocutor to interlocutor, on the other hand was deemed to have no interpretative value in the context of alignment for FL acquisition and was not considered during coding. For the same reason, instances where the dialog system repeated an n-gram that was used by the learner within the five-message threshold was automatically coded as *interlocutor to learner*, even if the n-gram also occurred in one of the dialog system's previous messages (see Example 3).

Example 3

Coding of lexical alignment from interlocutor to learner with previous occurrence in dialog system's production

28	System	Please could you confirm if you would like <u>to book room Alpha for the 12 octobre 2019?</u>	<i>Unprimed</i>
29	Learner	Yes I confirm <u>to book room Alpha for the 12 octobre 2019</u>	<i>Learner to interlocutor</i>
30	System	Please could you confirm if you would like <u>to book room Alpha for the 12 octobre 2019?</u>	<i>Unprimed</i>

Coding of Structural Alignment

Following the procedure described in Dao et al. (2018) and Michel and Cappellini (2019), structural alignment was identified by initially identifying and coding frequent grammatical structures. Frequent grammatical structures were defined as grammatical forms that occurred at least five times in the entire corpus. The authors followed this procedure based on Goldberg (2006), who argued that “structures need to occur with sufficient frequency to be considered as an independent construction” (Dao et al., 2018, p. 304).

Frequent structures were searched top-down based on the lists reported in Dao et al. (2018) and Michel and Cappellini (2019) and complemented with structures found bottom-up in the data. This procedure yielded a total of eight different grammatical structures that are listed along with examples in Table 6. Inclusion criteria involved considering whether the structure in question was relevant to the EFL learning and teaching context, i.e., were typically taught in grammatical instruction in EFL classrooms.

Table 6

List of Frequent Syntactical Structures for Coding of Structural Alignment

Structure	Definition	Examples
Copula <i>be</i>	The use of copula <i>be</i>	All of the three rooms <u>are</u> equipped for presentations. yes it's correct
<i>Have</i> + noun phrase	The use of <i>have</i> with a noun phrase as complement	Room Alpha <u>has no natural light</u> . Which room <u>has the most place to have conversations</u>
<i>For</i> + noun phrase	The use of <i>for</i> with a noun phrase as complement	Room Alpha provides space <u>for up to 270 people</u> . and i need some tables <u>for making groups</u> Is there anything else I can do <u>for you</u> ?
Modal verb + infinitive	The use of a verb phrase consisting of a modal verb (i.e., an auxiliary verb that expresses modality such as <i>can/could, shall/should, may/might, will/would, must</i>) as its head followed by an infinitive	Please <u>could</u> you <u>give</u> me your name? No, I <u>would like</u> to book a room for 1500 peoples The room <u>should be</u> able to fit up to 150 people.
<i>To</i> -infinitive	The use of <i>to</i> -infinitive, i.e., an infinitive verb preceded by the infinitival particle <i>to</i>	We need also a place <u>to speak</u> in little groups. it must be posible <u>to make</u> a presentation i want <u>to book</u> the room gamma
Existential <i>there</i>	The use of the dummy pronoun <i>there</i> as subject	Is <u>there</u> anything else I can do for you? Is <u>there</u> a good atmosphere? Are <u>there</u> any restaurants at the hotel?

Structure	Definition	Examples
Question with subject-auxiliary inversion	The use of an interrogative clause that is marked by subject-auxiliary inversion	<u>Does the room beta</u> have a good atmosphere? <u>Could you</u> please say it again?
Wh-interrogative	The use of an interrogative clause that is marked by the presence of an interrogative word, i.e., <i>who</i> , <i>what</i> , <i>when</i> , <i>where</i> , <i>why</i> , <i>how</i> , <i>which</i>	<u>How</u> may I assist you today? Yes <u>how</u> many would it cost? <u>Which</u> room have the best atmosphere?

Coding of structural alignment was done in two steps: First, the eight frequent structures were searched in each script and assigned the main code *structural alignment*, given that they occurred at least five times within the respective interaction (Dao et al., 2018; Michel & Cappellini, 2019) and in both the interlocutor's and the learners' production. This preliminary coding was done by the author.

Once the structures had been identified in the scripts, each segment pre-coded as *structural alignment* was assigned a subcode as a second step, indicating the direction of alignment or the subcode *unprimed*, respectively, if not preceded by the same structure within five messages. The coding followed the same procedures and used the same subcodes indicating direction of alignment as the coding for lexical alignment described above, with the exception of the *with error/error-free* distinction for the subcode *interlocutor to learner*, which was not relevant for the coding of structural alignment. So, as in lexical alignment, structural alignment was classified into the subcodes *learner to learner (self)*, *learner to interlocutor*, *learner to both*, or *interlocutor to learner*. Again, the direction from interlocutor to interlocutor was not considered and the interlocutor's reuse of a structure that was previously used within five messages by the learner as well as the interlocutor, was coded as alignment from *interlocutor to learner* rather than from *interlocutor to both*.

The following examples illustrate the coding procedure for the different directions of structural alignment for various structures. As will be seen in the examples, highly frequent structures with dense use in a dialog such as copula *be* often resulted in multiple alignment and overlap of different types of alignment within the same sequence. Thus, in Example 4, the learner in message 19 reuses the *wh*-question structure from the interlocutor's production in message 18. In message 24, i.e., five

messages further, the learner reuses the same structure, aligning to their own production from message 18. The use in message 24 was coded as *learner to learner* alignment rather than *learner to interlocutor* or *learner to both* alignment since it occurs eight lines from the previous interlocutor use, i.e., outside the five-message threshold.

Example 4

Coding of structural alignment from interlocutor to learner despite previous occurrence in dialog system's production

18	System	For <u>what</u> date would you like to reserve room Alpha?	<i>Unprimed</i>
19	Learner	but <u>how</u> much is it?	<i>Learner to interlocutor</i>
20 – 23		[...]	
24	Learner	<u>what</u> about the atmosphere in this room	<i>Learner to interlocutor</i>

Example 5 contains multiple alignment categories. The learner in message 4 reuses the structure *for* + noun phrase from the interlocutor's production in message 2. In message 6, the learner reuses the same structure again, thus aligning to their own as well as the interlocutor's production. The same structure is reused again by the interlocutor in messages 8, 9, 10, and 11. Since these occurrences are within five messages from message 6 where the learner previously used the structure, they were coded as *interlocutor to learner* alignment.

Example 5

Coding of multiple directions of structural alignment

2	System	Thank you <u>for</u> contacting One King Hotel. I will gladly assist you in finding a room.	<i>Unprimed</i>
3	System	How may I assist you today?	
4	Learner	I am searching <u>for</u> a room in your hotel	<i>Learner to interlocutor</i>
5	System	We have multiple rooms available.	<i>Learner to learner</i>

6	Learner	So the room i'm searching for should be big enough <u>for</u> minimum 150 people	<i>Learner to both</i>
7	System	Yes, certainly.	
8	System	There are three rooms available <u>for</u> 150 people.	<i>Interlocutor to learner</i>
9	System	Room Alpha provides space <u>for</u> up to 270 people.	<i>Interlocutor to learner</i>
10	System	The Beta room has space <u>for</u> a total of 190 people, indoor and outdoor.	<i>Interlocutor to learner</i>
11	System	There is space <u>for</u> 200 people in room Gamma.	<i>Interlocutor to learner</i>

Similarly, in Example 6, the learner in message 4 reuses copula *be* from the interlocutor's production in message 2, and again in message 7 as *learner to both* alignment. The learner reuses the same structure again in message 11. Since this occurrence is within five messages from the learner's own previous use in messages 4 and 7, but outside the threshold of the interlocutor's previous use in message 2, it was coded as *learner to learner* alignment. The structure is then reused by the interlocutor in message 12, aligning to the learner's production in messages 7 and 11. The next occurrence is produced by the learner in message 13, so within five messages of a previous learner as well as interlocutor production, thus considered *learner to both* alignment. The interlocutor reuses the structure once more in message 16, aligning to the learner's use in messages 11 and 13. Note that although the structure was equally used by the interlocutor within the five preceding messages, the occurrence in message 16 was coded as *interlocutor to learner* since alignment from *interlocutor to interlocutor* was not considered for coding.

Example 6

Coding of multiple directions of structural alignment

2	System	Thank you for contacting One King Hotel. I <u>am</u> happy to help you find a room for your event.	<i>Unprimed</i>
3	System	How may I help you?	

4	Learner	Hello i'm <u>Dariana</u> from the company of a marketing. I have question. I need a room for 150 people for a presentation with a great area. And also I want more Information of you for a good room for our presentation.	<i>Learner to interlocutor</i>
5 – 6		[...]	
7	Learner	My Name <u>is</u> Dariana.	<i>Learner to both</i>
8 – 10		[...]	
11	Learner	That's great. And how many <u>is</u> it?	<i>Learner to learner (self)</i>
12	System	I am afraid I have trouble understanding. Please could you rephrase?	<i>Interlocutor to interlocutor</i>
13	Learner	How many will <u>be</u> costs?	<i>Learner to both</i>
14 – 15		[...]	
16	System	What would <u>be</u> the date for the booking?	<i>Interlocutor to interlocutor</i>

Example 7 similarly shows multiple instances of *interlocutor to learner* alignment in messages 13, 14, and 15. The next occurrence is produced by the learner within five messages of the previous interlocutor's but outside the threshold of their own production, so as *learner to interlocutor* alignment. This is followed by two more occurrences of *interlocutor to learner* alignment in messages 20 and 21.

Example 7

Coding of multiple directions of structural alignment

12	Learner	There are some more critereas i <u>would</u> like to say	<i>Unprimed</i>
13	System	I am sorry, I did not understand that. <u>Could</u> you please try one more time?	<i>Interlocutor to learner</i>
14	Learner	There are some more points	
15	System	Please <u>could</u> you tell me who I am chatting with?	<i>Interlocutor to learner</i>
16	Learner	This is Mithusha	
17	System	I am sorry, I did not understand that. <u>Could</u> you please try one more time?	<i>Interlocutor to learner</i>

18	Learner	Yes sure, I <u>would</u> like to book a room for minimum 150 people in your hotel. The room <u>should</u> let us do a presentation and also meeting for small groups. It <u>should</u> also have a good atmosphere. It <u>should</u> make the guests happy. Do you have room for maximum 1500 Franks?	<i>Learner to interlocutor</i>
19	System	Yes, certainly.	
20	Learner	Thank you for contacting One King Hotel. I <u>will</u> gladly assist you in finding a room.	<i>Interlocutor to learner</i>
21	System	Please <u>could</u> you tell me who I am chatting with?	<i>Interlocutor to learner</i>

As in lexical alignment, there also occurred unprimed reuse of structures, i.e., the reuse of a structure beyond the five-message threshold. This is illustrated in the examples below. In Example 8, the learner reuses *to*-infinitive in message 12 after using it previously in message 1, i.e., outside the five-message threshold.

Example 8

Coding of unprimed reuse of structures

1	Learner	Hello, i am searching for a room suitable for our company event that costs 1'500 Francs maximum. It will take place on 12th October 2019. The room should be big enough <u>to</u> fit a minimum of 150 people. We also need to be able to play video and audio in the room.	<i>Unprimed [first occurrence]</i>
2 – 11		[...]	
12	Learner	User: Will we be able <u>to</u> play video and audio in room alpha?	<i>Learner unprimed</i>

Similarly, in Example 9, the learner in message 20 reuses a structure with subject-auxiliary inversion from the interlocutor's message 7, i.e., outside the five-message threshold. The next occurrence in 26 is outside the threshold again. The occurrence in message 35 is equally outside the threshold.

Example 9*Coding of unprimed reuse of structures*

7	System	Excuse me, <u>could you</u> rephrase this please?	<i>Unprimed [first occurrence]</i>
8 – 19		[...]	
20	Learner	<u>are there</u> monitors and speakers	<i>Learner unprimed</i>
21 – 25		[...]	
26	Learner	<u>Have you</u> got a chance to do an presentation in the Alpha room?	<i>Learner unprimed</i>
27 – 34		[...]	
35	System	So sorry, I could not understand your last message. <u>Could you</u> please say it again?	<i>Interlocutor unprimed</i>

Measures for Complexity, Accuracy, and Communicative Adequacy

To answer research question 2 on the differences in learners' task performance between learner-system interaction and SCMC, the learners' portion of the interactions was further analyzed in terms of complexity, accuracy, and communicative adequacy. The following sections present the measures that were used to assess these three aspects of learner production along with the procedure to calculate them. A preceding section explains the procedure of the preliminary coding of *errors*, *AS-units*, and *requests* that was necessary to derive the elements for the calculation of the measures.

Preliminary Coding of Errors, AS-Units, and Requests

As will be explained in more detail below, the calculation of the complexity and accuracy measures relied on the preliminary coding of *errors*, *AS-units*, and *requests* in the interaction scripts. The code system for the coding of these elements was established based on a separate sample of learner-system interactions from a pilot study, as in the coding of interactional features (see section Coding of Interactional Features). Initial codes were developed bottom-up based on the learner-system scripts, as will be described in more detail in the following sections. The resulting code system was then applied to the SCMC scripts to verify that it was applicable to the SCMC data.

As a result of this verification process, the following adaptation to the coding guidelines for *requests* was done, as already mentioned above: Comparing the request structures in the two interaction modes, learners in SCMC frequently split their utterances including requests over multiple messages. When interacting with the dialog system, however, this was prevented by design since the dialog system responded to each message immediately after it was sent by the user. Therefore, the coding guidelines for the code *requests* were complemented with a specification that accounted for this kind of message structuring behavior. All final coding definitions, guidelines, and examples can be found in the codebook for complexity, accuracy, and communicative adequacy (for link, see Appendix H).

Coding of Errors

As discussed in the literature review, the notion of *error* in EFL teaching and learning raises the question of what constitutes the norm in language use as the normative standard may vary between, e.g., different modalities (written/oral) or dialects. In this study, the criteria for identifying errors in the data were set in line with the context of the study, so tuning them to Standard English norms, which is the language norm of EFL teaching and learning in schools. However, erroneous punctuation and capitalization were not coded as errors in the analysis as they were expected to be frequently neglected in the chat context, which would have disproportionately affected the accuracy measures.

Another question pointed out in the literature review is whether the severity of an error should be taken into account, as some errors may be so grave that they impair communication. Since in the present study, success of communication is evaluated separately by the communicative adequacy measure, errors were not weighted for severity. Similarly, deviations from pragmatic norms were excluded from the error coding since pragmatic appropriateness was covered through the rating of request appropriateness (see section Accuracy Measures).

Based on these considerations, *errors* were operationalized as grammatical, lexical, and orthographic (excluding punctuation and capitalization) deviations from Standard English. To allow for a more fine-grained analysis of linguistic accuracy and

to establish a basis for the coding of the interactional feature *recasts* and its subcodes (see section Coding of Recasts), the identified errors were further classified as *grammar*, *lexis*, or *spelling*. The following paragraphs present the guidelines that were followed for the coding of each error type. Detailed coding guidelines and examples can be found in the codebook (for link, see Appendix H).

- The code *grammar* was assigned to deviations from morphological or syntactical Standard English norms relating to the word structure such as incorrect plural forms (e.g., “There are some more critereas”), incorrect verb number/tense/aspect/mood (e.g., “and what about showing videos? is that working?”), omission of third person singular -s (e.g., “Which room have the best atmosphere?”) as well as structures at the phrase and clause level such as omission of words (e.g., “The should have space for at least 150 peoples”) superfluous words (e.g., “Do they have a beamer in?”), incorrect use of *a/an* (e.g., “I need room for 150 Persons”), word misorder (e.g., “In which room can i make also private speak”), or omission of subject-verb inversion or dummy *do* in interrogative sentences (e.g., “How much space has it?”).
- The code *spelling* was assigned to orthographic errors including errors that were likely the result of typing mistakes (e.g., “what about the atmosphere in thisroom”). As mentioned above, punctuation errors such as omission of end-of-sentence period (e.g., “What's the price for room Gamma”), omission of apostrophe (e.g., “im interesed at the room Gamma”) as well as capitalization errors such as omission of capitalization at the beginning of a turn (e.g., “yes but we are 150 persones”) or omission of capitalization of names (e.g., “Ok, we'll go with room beta.”) were not considered.
- The code *lexis* included formal errors such as the misselection of suffixes (e.g., “i want to reservate a room”) or prepositions (e.g., “we would like to come at the 12 june”) or pronouns (e.g., “I'm searching for a room which [...]”), invention of non-existing words (e.g., “Is it like very extra?”) as well as semantic errors such as the choice of a lexical item in a context where its meaning is inappropriate or inaccurate (e.g., “it must be posible to make a presentation”), and the use of German (e.g., “and what about having apero?”).

As outlined in more detail in the codebook (for link, see Appendix H), a word or a group of words could be assigned more than one error subcode (e.g., “Which room is the best for smalltalks [...]” [grammar error and spelling error]). However, when a word contained a lexical as well as grammatical and/or spelling error, only the code *lexis* was assigned (e.g., “Can we made a Presentation [...]”) [lexis error; grammar error not coded separately]). A word containing multiple errors of the same type were assigned the respective code once. When a learner made the same error repeatedly within the interaction, it was coded each time.

Coding of AS-Units

AS-units were operationalized following Foster et al. (2000) who defined an AS-unit as a “single speaker’s utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either” (p. 365). Following the rationale by Foster et al. (2000), AS-units were chosen as unit of analysis in the present study as it allows for the analysis of units consisting of “multi-clause units” as well as “sub-clausal units” common in spoken language. “Sub-clausal units” are units consisting of “either one or more phrases which can be elaborated to a full clause by means of recovery of ellipped elements from the context of the discourse or situation [...] or a minor utterance”, i.e., “[i]rregular sentences’ or [n]onsentences” such as “Thank you very much” or “Yes” (p. 366). Since written SCMC combines characteristics of both oral and written language (Adams et al., 2015), the AS-unit has been used as a more valid unit of analysis in studies to evaluate the accuracy and complexity of learners’ task performance in written SCMC (e.g., Adams et al., 2015; Hsu, 2017; S. Sauro, 2012).

For coding, punctuation and message structuring were considered when establishing the boundaries of AS-units, following the suggestion by Adams et al. (2015). So, the boundaries of an AS-unit were generally considered to be marked by punctuation (i.e., comma, period) or a new message. Similarly, coordinated elements such as phrases or clauses connected with a coordinator (i.e., *and*, *or*, *but*) were considered part of the same AS-unit unless they were separated by punctuation or a new message. These coding rules are illustrated in Example 10. Further examples and

detailed coding guidelines can be found in the codebook (for link, see Appendix H). When identifying AS-units, they were additionally categorized as *error-free* or *with error* based on the preliminary error coding (see section Coding of Errors) as a basis for accuracy measurement.

Example 10

Coding of AS-units

Yes [AS-unit 1] sure, [AS-unit 2] I would like to book a room for minimum 150 people in your hotel. [AS-unit 3] The room should let us do a presentation and also meeting for small groups. [AS-unit 4] It should also have a good atmosphere. [AS-unit 5] It should make the guests happy. [AS-unit 6] Do you have room for maximum 1500 Franks? [AS-unit 7]

Coding of Requests

As will be explained in more detail below, requests were chosen as unit of analysis to measure the pragmatic appropriateness of the learners' production. Since only the pragmatic appropriateness of the learner's (not the dialog system's or peer's) was of interest, only requests from the learner's portion of the interaction were considered for analysis. Consequently, and following the definition commonly used in L2 and FL research that is based on the conception of requests by Searle (Timpe-Laughlin & Dombi, 2020), the present study defined requests as directive speech acts that have the function to get the interlocutor to do something for the learner. Following the Cross-Cultural Study of Speech Act (CCSARP) framework proposed by Blum-Kulka et al. (1989), requests were furthermore conceptualized as consisting of a head act, i.e., the "minimal unit which can realize a request" (p. 275) and support moves that occur either within the head act (i.e., internal modifiers) or before or after the head act (i.e., external modifiers) and typically serve to mitigate the illocutionary force of the head act.

Requests were coded in the learners' production following this definition and in a two-step procedure: First, requests were identified and coded including head act and modifiers and the coding was discussed and matched with the second rater. Example 11, Example 12, and Example 13 illustrate this preliminary assigning of the main code

request. As seen in the examples, a learner message may contain a single request (Example 11 and Example 12). When a message contained requests on different topics, each request was coded individually (Example 13). As also seen, the coding included both requests for action (Example 12) and requests for information (Example 11). Further examples and detailed coding guidelines can be found in the codebook (for link, see Appendix H).

Example 11

Assigning of main code Request

Okey and witch room is conact with video and electric things? [request]

Example 12

Assigning of main code Request

Ok, we'll go with room beta. Thanks for your advise. [request]

Example 13

Assigning of main code Request

Not quite sure. The room should be able to fit up to 150 people. [request 1] We need a big screen so every one of said 150 people is able to see the presentation our "higher ups" made. [request 2] Our workers should have the ability to chat with each other in small groups. [request 3] We need the room to give us a nice and fun atmosphere. [request 4] It should make abig impression on our guests, but shouldn't come off as "too exaggerated". [request 5] We're willing to pay 1'500 under these conditions. [no request] Is there a room you could suggest for us? [request 6]

As a second step, the request components (i.e., head act, modifiers) were further classified into subcodes. This classification served to identify the strategy type used to realize the head act as well as the linguistic elements used to mitigate the illocutionary force of the request either within the head act (i.e., internal modifier) or outside the head act (i.e., external modifier). The subcodes were created based on previously developed taxonomies (Blum-Kulka et al., 1989; Cunningham, 2016a; Felix-Brasdefer, 2015). Where necessary, the definitions from the literature were adapted slightly to match the data and additional subcodes were added. An overview of the final main and subcodes is presented in Table 7 along with the respective sources.

Detailed code descriptions and further examples can be found in the codebook (for link, see Appendix H).

Table 7*Overview of Main and Subcodes for Requests*

Main code	Subcode	Definition	Example
Request strategy	Imperative (Cunningham, 2016a)	The grammatical mood of the verb in the utterance marks its illocutionary force as a request.	Please give my a answer to my question.
	Direct question (adapted from Cunningham, 2016a)	Typically a request for information that is conveyed by a direct question.	is it possible to communicate in small groups?
	Elliptical question	Typically a request for information that is conveyed by an elliptical question, i.e., an interrogative that lacks a verb and/or subject which can be recovered from the context.	And the Beta Room?
	Indirect question (adapted from Cunningham, 2016a)	Typically a request for information that is conveyed by an indirect question, i.e., a question conveyed by a subordinate interrogative.	I asked, if this is the "One King Hotel" in Toronto.
	Want statement (adapted from Cunningham, 2016a)	The learner expresses his or her desire that the action expressed in the request be carried out by the interlocutor.	i would like to reserve the room Alpha
	Assertion	The learner expresses his or her reservation request as a declarative statement.	Very nice, I book the room Gamma for the october 12th
	Reference to preparatory conditions (Blum-Kulka et al., 1989)	The utterance contains reference to a preparatory condition for the feasibility of the request such as the interlocutor's ability or willingness or the possibility to perform the action.	can you tell me about the Alpha, Beta, Gamma rooms

Main code	Subcode	Definition	Example
	Strong hint (Blum-Kulka et al., 1989)	A statement that contains implicit reference to the action without explicitly stating the intended action.	I am very intresed in the room Alpha
Internal modifier	Politeness marker (Cunningham, 2016a)	An optional element added to the head act to bid for cooperative behavior.	Could you <u>please</u> tell me prices of the rooms?
	Modal preterite (adapted from Blum-Kulka et al., 1989)	The verb of the head act is modified by a modal in the preterite form.	We <u>would</u> like to book the room Alpha
External modifier	Alerter (Cunningham, 2016a)	A linguistic device preceding the head act that is used to get the interlocutor's attention.	<u>Hello</u> , is this the One King Hotel in Toronto?
	Grounder (adapted from Blum-Kulka et al., 1989; Cunningham, 2016a)	An utterance in which the learner states his or her reasons or justifications for making the request.	<u>I have questions about your hotel</u> . Did you have a room for 150 persons?
	Agreer	The learner expresses that he or she is satisfied with the offering.	<u>That sounds great</u> . Is the room gamma fancy or like a normal room
	Appreciator (adapted from Cunningham, 2016a)	The learner expresses his or her gratitude for the (anticipated) performance of the action.	<u>Thanks</u> and how is the atmosphere in the Room Gamma?
	Disagreer	The learner expresses that he or she is dissatisfied.	System: Would you like me to book room None for the None? Learner: <u>no</u> , which room is the best for have also place for tidy conversation

Note. References indicate the source of the respective subcode. Subcodes without indication of source were developed anew.

Complexity Measures

The following sections present the measures that were used to investigate differences in the complexity and accuracy of learners' task performance between the two interaction types (system, SCMC). The measures were selected based on the following general considerations: First, given the exploratory nature of the analysis, they should include a range of measures to tap into different aspects of linguistic complexity and accuracy. Second, they should include those measures that had been shown to be affected by the learner-system interaction type (e.g., lexical and syntactic diversity, pragmatic appropriateness). Finally, they should include measures that are commonly used in EFL research to allow for comparability between studies.

Lexical Diversity Measured by MTL D

Lexical diversity refers to the range of different words that are used in the production, with a greater range indicating higher lexical diversity. Lexical diversity is a widely used complexity measure in L2 and FL research. The best-known lexical diversity index is the type-token ratio (TTR) (Templin, 1957), which is computed by dividing the number of different lexical items (types) by the number of total words (tokens) in a text. However, as TTR is sensitive to text length, alternative measures have been developed. Two of the most widely used alternative measures are the diversity index D (Malvern et al., 2004) and the Measure of Textual Lexical Diversity (MTLD) (McCarthy, 2005). For this study, MTLD was selected as it has been shown to be more robust with regard to text length variations than D (Koizumi, 2012; McCarthy & Jarvis, 2010) and text samples of various lengths containing between 100 and 2000 tokens (Crossley et al., 2009), two criteria that apply to the interactions in this study. In addition, Koizumi (2012) found text length sensitivity of MTLD in short texts of less than 100 tokens, with the effect being more pronounced in texts with high token differences (2012). The author therefore recommends that MTLD be used only for texts with at least 100 tokens and that have, if possible, a maximum of a 50-token difference between texts.

The lexical diversity index MTLD is "calculated as the mean length of sequential word strings in a text that maintain a given TTR value", the value being .720

(McCarthy, 2005, p. 384). In order to do the calculation, MTLN evaluates the TTR of each word of a text until the TTR factor size value .720 is reached. At this point, the TTR measurement is “reset” (McCarthy, 2005, p. 384) and starts again. Then, the total number of words in the text is divided by the total number of times the factor size value is reached. A second MTLN value is calculated in the same way, processing the text in reverse from the last to the first word. The final MTLN value is the mean of the two values (see Koizumi [2012] or McCarthy [2005] for a detailed description of the MTLN calculation).

In the present study, the learner’s portion of each interaction script was first spellchecked and subsequently entered into the web tool Coh-Metrix 3.0 (D. McNamara et al., 2003) to obtain the MTLN score. As mentioned above, it has been recommended to use MTLN with texts of 100 tokens or more only due to its sensitivity to text length. However, since both sets of scripts in this study (system, SCMC) contained an equal number of texts with fewer than 100 tokens, shorter texts were expected to impact the score equally in both groups and were consequently not excluded from the analysis.

Overall Syntactic Complexity Measured by Number of Words per AS-Unit

Overall syntactic complexity was assessed by measuring the length of AS-units in terms of number of words per AS-unit. AS-unit length is a commonly used measure to gauge the overall syntactic complexity of L2 and FL learner production (Norris & Ortega, 2009). Increased length of AS-units indicates increased complexity, as AS-units may become longer, for instance, through the coordination and subordination of clauses as well as through the addition of elements to the constituent clauses or phrases such as complements (e.g., relative clauses, verb + *to*-clauses) and modifiers (e.g., attributive adjectives, adverbials) (Bulté & Roothoof, 2020; Norris & Ortega, 2009).

In the present study, the measure was calculated by dividing the total number of words by the total number of AS-units each learner produced. Total number of words was measured in the spellchecked learner texts using Coh-Metrix (D.

McNamara et al., 2003). AS-units were coded by hand following the procedure described above (see section Coding of AS-Units in the chapter Methods).

Syntactic Similarity Measured by SYNSTRUTt

The SYNSTRUTt index provides a value that indicates the syntactic similarity between all sentences in a text. A lower value indicates lower syntactic uniformity, and hence, greater syntactic diversity. To compute the value, SYNSTRUTt finds the largest common syntax subtree between two sentences, combining all sentences across a text. It then divides the number of tree nodes in the common subtree by the sum of the nodes in both sentences' trees minus the common nodes. For instance, to compute the SYNSTRUTt score for the two sentences 1) "The man came." and 2) "He entered the door.", SYNSTRUTt would analyze six common nodes (e.g., the noun phrases "The man" in sentence 1 and "He" in sentence 2; the verb phrases "came" in sentence 1 and "entered the door" in sentence 2) and 12 different nodes (e.g., noun phrase with determiner "The man" in sentence 1 vs. pronoun as noun phrase "He" in sentence 2; no complement in sentence 1 vs. verb complement "the door" in sentence 2), resulting in the SYNSTRUTt score .50 (McNamara et al., 2014). SYNSTRUTt thus measures the uniformity and consistency of syntactic structures at the level of the clause, phrase, and part of speech. A more detailed description of the calculation of SYNSTRUTt may be found in McNamara et al. (2014).

In the present study, the SYNSTRUTt score for each learner production was obtained from Coh-Metrix (D. McNamara et al., 2003).

Accuracy Measures

As pointed out in the literature review above (see section Complexity, accuracy, fluency), the accuracy of learner production may be gauged by means of general measures (e.g., number of error free clauses) as well as specific measures (e.g., target-like use of past forms) relating to task-specific aspects of accuracy. The two general measures *errors per AS-unit* and *error-free AS-units per AS-unit* were selected to assess the overall linguistic, i.e., lexico-grammatical and orthographic accuracy, of the production.

Errors per AS-Unit

As explained above (Coding of AS-Units), the current study used AS-units as syntactic unit to analyze the learner-system and SCMC interaction data. Errors per AS-unit is a measure of accuracy based on this unit of analysis commonly used in the context of L2 and FL FTF and SCMC interaction. In the present study, the measure was calculated by dividing the total number of errors by the total number of AS-units produced by each learner. The number of errors was derived from coding, which included the coding of errors in terms of grammatical, lexical, and orthographic (excluding punctuation and capitalization) deviations from Standard English (see section Coding of Errors). The number of AS-units was also derived from coding (see section Coding of AS-Units).

Error-free AS-Units per AS-Unit

As an additional common measure of L2 and FL production accuracy, the ratio of error-free AS-units per AS-unit was calculated by dividing the number of error-free AS-units by the total number of AS-units each learner produced. The number of error-free AS-units was derived from coding, which included the coding of AS-units as *error-free* or *with error* (see section Coding of AS-Units). This categorization was based on the preliminary coding of *errors* (see section Coding of Errors).

Appropriateness of Requests Measured by Mean Request Appropriateness Score and Number of Inappropriate Requests per Request

The appropriateness of learner requests was included as a task-specific measure of accuracy. It was deemed relevant to include an accuracy measure relating to requests, for several reasons: First, requests were particularly pertinent to the task used in this study, which essentially requires the learner to produce a series of requests (e.g., information requests, booking requests). Second, requests and their function in communication and social interaction are highly relevant for EFL learning and teaching: Requests constitute an inherently face-threatening type of speech act as they function to get the interlocutor to do something in the interest of the speaker and learning to produce adequate requests thus plays a crucial role for the development of

pragmatic competence in L2 and FL learning (Ellis, 2008, p. 173). For this reason, requests are commonly analyzed in studies to evaluate the pragmatic appropriateness of learner production (Felix-Brasdefer, 2019), including in studies on SCMC (e.g., Cunningham, 2016a) and dialog systems (Sydorenko et al., 2019; Timpe-Laughlin & Dombi, 2020).

To evaluate the appropriateness of learner requests in the present study, first, each request was rated on a three-level scale ranging from *inappropriate* (corresponding to the score value 1), through *partly appropriate* (score 2), to *fully appropriate* (score 3). Each level corresponded to a code in MAXQDA 2020 which contained descriptors pertaining to the appropriateness of the request in terms of its degree of directness, the adequacy of its internal and external modification, and its grammatical and discourse appropriateness. The descriptors were adapted from the taxonomy proposed in the study of Cunningham (2016b) on request appropriateness in FL telecollaboration. The detailed coding guidelines along with examples can be found in the codebook (for link, see Appendix H). Coding was conducted by two coders to ensure intercoder reliability, which will be described in detail below (see section Intercoder Reliability). Based on the ratings of each individual request, the following two measures were calculated:

- Mean request appropriateness score: The mean request appropriateness score was calculated by adding up the scores of each individual request and dividing the sum by the total number of requests the learner produced.
- Inappropriate requests per request: For each learner, the total number of *inappropriate* ratings were divided by the total number of requests produced.

Overview of Complexity and Accuracy Measures

An overview of the measures of linguistic complexity and accuracy described in the preceding paragraphs is shown in Table 8. As mentioned at the beginning of this chapter, only the learners' portions of the interaction were used when calculating the measures.

Table 8
Overview of Measures of Complexity and Accuracy

Focus	Measure	Calculation
Complexity		
Lexical diversity	Measure of Textual and Lexical Diversity (MTLD)	Obtained from Coh-Metrix: mean length of sequential word strings in a text that maintain a TTR value of 0.720
Overall syntactic complexity	Words per AS-unit	Total number of words divided by the total number of AS-units
Syntactic similarity	SYNSTRUTt index	Obtained from Coh-Metrix: combining all sentences across the text, SYNSTRUTt finds the largest common subtree between two sentences, then divides the number of nodes in the common subtree by the sum of the nodes in both sentences' trees minus the common nodes
Accuracy		
Overall lexico-grammatical and orthographic accuracy	Errors per AS-unit	Total number of errors divided by total number of AS-units
	Error-free AS-units	Number of error-free AS-units divided by total number of AS-units
Appropriateness of requests	Mean request appropriateness score	Sum of the individual request appropriateness rating scores divided by the total number of requests
	Inappropriate requests per total requests	Number of inappropriate requests divided by total number of requests

Note. The measures were calculated for each learner using only the learners' portion of the interaction.

Communicative Adequacy Measure

Evaluating the communicative adequacy of the learners' performance was deemed particularly relevant in the context of this study. As discussed in the literature review (see section Task-based Interaction), the evaluation of communicative adequacy specifically reveals the pragmatic and sociolinguistic competences

underlying the performance and thus indicates in how far the task enables the learner to engage with these two components of communicative competence as intended. The level of engagement with all three aspects of communicative competence (i.e., linguistic, sociolinguistic, pragmatic), in turn, is important to evaluate as it is stated as the goal of the participants' commercial VET school EFL curriculum, as described in the introduction. Moreover, it was expected that the particular 'social' context of learners interacting with an automated agent rather than another human would impact the learners' performance with respect to its sociolinguistic and pragmatic dimensions.

As far as possible, this study relied on previous definitions and operationalizations when assessing the communicative adequacy of the learners' production to allow comparability across studies. As pointed out in the Literature Review (see section Communicative Adequacy Measure), no agreement exists in the literature on the definition and assessment of communicative adequacy (Kuiken et al., 2010b). So, synthesizing existing definitions, this study used the definition presented in the literature review, conceptualizing communicative adequacy as the learner's ability to achieve the task's goals by using language that is adequate to the specific communicative context.

According to this definition, communicative adequacy was evaluated along the two dimensions task completion and appropriateness to the communicative context. This analytical rather than holistic assessment was chosen since the analysis should not only yield an overall score but also provide qualitative information on the individual dimensions of communicative adequacy.

Task completion was operationalized as the degree to which the learners completed the task according to the prompt, i.e., booked one of the three rooms after requesting the required information about at least one of the rooms. Appropriateness to the communicative context was operationalized in terms of genre appropriateness, i.e., the adherence of the learners' discourse to the genre evoked by the task. Assessing the production's appropriateness to the communicative context by means of evaluating its genre appropriateness was deemed a suitable procedure since genres can be understood to represent communicative context (Rijlaarsdam et al., 2006) in that they are the conventionalized discourse practices that arise in response to recurrent

communicative situations (Hyland, 2018). The teaching of genre and genre awareness is, moreover, a central objective of English teaching (Melissourgou & Frantzi, 2017; Paltridge, 2014). This becomes evident in the commercial VET EFL curriculum which requires receptive and productive skills in a variety of specific genres such as small talk, reservation, confirmation of reception, or job application and so on. To assess the genre appropriateness of the learners' performance, the production of each learner was assessed in terms of generic structure and generic style, which are the two major components that make up a particular genre and are commonly explored in genre analysis for English teaching (Hyon, 2018).

To conduct the rating of these three aspects (task completion, generic structure, and generic style) a code for each aspect was added to MAXQDA 2020 (VERBI Software, 2020). Each code contained three subcodes pertaining to three rating levels and corresponding numerical scores, as shown in Table 9.

Table 9

Subcodes/Levels and Corresponding Scores for the Rating of Communicative Adequacy

Aspect	Subcode/Level	Score
Task completion	Not successful	1
	Partly successful	2
	Successful	3
Generic structure	Inappropriate	1
	Partly genre-appropriate	2
	Fully genre-appropriate	3
Generic style	Informal	1
	Mixed	2
	Formal	3

The rating criteria for each subcode/level were specified in the code memo. The rating criteria were established based on existing communicative adequacy rating scales (De Jong et al., 2012; Hulstijn et al., 2012; Kuiken & Vedder, 2018; Revesz et al., 2016) and considering the participants' coursebook, in order to match the assessment to the learners' previous instruction and curriculum. Thus, the *generic style* criteria were established based on the coursebook's guidelines and sample texts for written

communication such as emails and letters specifying the linguistic forms that are “more formal” (e.g., complete sentences, use of punctuation, formal address etc.) or “less formal” (e.g., contractions, abbreviations, ellipses etc.) (Whitby, 2013). The “more formal” style was chosen as benchmark for the highest rating. A list with the style features as specified in the coursebook was added to the rating criteria in the code memos. The “more formal” style, as the coursebook specifies, is suitable for communicating with someone outside the company with whom learners are not familiar, which describes the context of the task used in the study. However, it must be noted that the coursebook’s guidelines were based on emails and letters only and did not consider chat or chatbot interaction, which, it could be argued, require a less formal style. So, the lack of rating criteria based on chat communication and more suitable to assessing stylistic appropriateness in the chat context needs to be considered as a limitation of this approach.

Similarly, the benchmark for a *fully genre-appropriate* rating of the generic structure was set based on an analysis of the service encounter genre, i.e., the genre evoked by the task used in this study, presented in Felix-Brasdefer (2015) as well as a genre analysis of the service encounter interactions in the participants’ coursebook. Based on this analysis, the following model representing the generic structure that the learner production needed to follow to achieve the highest rating was developed and added to the code memos:

1. Greeting
2. Self-presentation (name, position) and/or presentation of reason for contact
3. Decision stage: information requests and provision
4. Reservation request
5. Reservation closure
6. Leave-taking

Again, the coursebook samples only included emails and letters, while the genre analysis by Felix-Brasdefer (2015) was based on FTF interactions, which limited the appropriateness of the rating criteria to some degree.

As a note, some existing rating scales for communicative adequacy include aspects of interactive adequacy (e.g., Revesz et al., 2016), i.e., the degree to which the

learner produces comprehensible language. In this study, however, this aspect was excluded since the dialog system's 'comprehension' was not merely impacted by how well-formed the learner's production was but also by technical factors, making an assessment based on the interlocutor's comprehension inadequate.

Intercoder Reliability

To ensure reliability of the code systems (interactional features; complexity, accuracy, communicative adequacy), the coding was conducted by two coders following the intercoder procedure suggested by Kuckartz and Rädiker (2019). The two coders were the author and a second coder, who is a Ph.D. student in EFL language teaching and learning and holds a master's degree in cognitive science with a specialization in computer linguistics.

First, as mentioned at the beginning of this chapter, initial code systems based on a separate sample of scripts from a pilot study were created by the author. These code systems were then presented and explained to the second coder. Subsequently, both the author and the second coder independently applied the code systems to four learner-system scripts, working line by line.

For ease of coding, the code systems were applied in several steps (Kuckartz & Rädiker, 2019, p. 270), as mentioned above: As a first step, *errors* were identified, applying the respective subcode (i.e., *grammar*, *spelling*, and *lexis*). As a next step, the codes for the complexity and accuracy measures were assigned: AS-units were coded directly at the subcode-level (i.e., as *error* or *error-free*), whereas requests were initially coded without assigning their subcodes. Similarly, for the interactional feature recasts only the main code *recast move* was assigned at this stage, and negotiation of meaning were identified as entities by initially coding the component *indicator*. Syntactic and lexical alignment were categorized directly into subcodes relating to the direction of alignment (i.e., *learner to interlocutor* etc.). As a final step, all remaining subcodes were assigned (i.e., for requests: type of strategies and modifiers; for recasts: linguistic focus, intensity, completeness etc.; for negotiation of meaning: trigger, response etc.). The communicative adequacy rating was also completed at this final stage of coding.

Intercoder agreement was checked after each coding step using the MAXQDA 2020 intercoder agreement function (segment comparison level) (VERBI Software, 2020). Any questions and uncertainties were discussed between the two coders and the coding guidelines were clarified and anchor examples were added. Where necessary, subcodes were merged or further subcodes were added. Any discrepancies in the code assignments were also discussed until agreement was reached (Kuckartz & Rädiker, 2019, p. 269). Once the coders reached agreement, the next stage of coding was completed. After completing all coding steps for the first set of scripts, another set of four learner-system scripts was coded independently by both coders, following the same procedure as described above. The author then coded the remaining eleven learner-system scripts. The second coder separately coded a representative 25 percent (four scripts) random selection and intercoder agreement was checked again, as detailed below.

The same procedure was repeated when coding the total 18 SCMC scripts: Both coders separately coded a set of four scripts and checked and discussed all disagreements. The author then coded the remaining 14 scripts of which the second coder coded a 25 percent (four scripts) random selection.

Intercoder agreement on the sample of a total of eight scripts (four learner-system, four SCMC) that were coded by both coders in the final coding step is reported in the table below. Due to the large number of subcodes in a relatively small double-coded sample, some subcodes only had a small number of occurrences in the sample. Nonetheless, to base the agreement measures of subcodes on a reasonable number of codings, intercoder agreement is documented for the overall agreement across all subcodes for a given category (see Table 10). The complete agreement measures for each code along with absolute occurrences can be found in the appendix (see Appendix I). As a further note, for codes where the segments to be coded were defined beforehand as part of the multi-step coding procedure described above (e.g., request subtypes, recast subtypes etc.), the agreement measure had to be corrected for chance. Therefore, the chance-corrected Brennan-Prediger (1981) kappa coefficient in addition to percentage agreement is also reported for codes that were applied to pre-defined segments (Kuckartz & Rädiker, 2019).

As can be seen in the table, percent agreement on all codes but the code *recast move* (see discussion in the next paragraph) that did not rely on pre-defined segments ranged from 79.07% to 100%, which was within or above the 70% – 80% that are commonly considered the minimum acceptable level (McHugh, 2012; Stemler, 2004). The Brennan-Prediger kappa values calculated to measure agreement for codes applied to pre-defined segments ranged from .79 to 1.00, which is considered moderate (.60 – .79) to almost perfect (above .90) agreement (McHugh, 2012).

With respect to the code *recast move*, which reached an agreement of 61.54%, it needs to be noted that the agreement value is likely skewed due to the low prevalence of *recast move* codings (total of 13 instances) in the sample selected for calculation of intercoder agreement. Notwithstanding, the following measures were taken to ensure reliability of the *recast move* code: All disagreements were discussed between the two coders until agreement was reached on all instances and the coding guidelines were clarified further. The first coder (author) subsequently reviewed all *recast move* codings in the entire sample based on the updated coding guidelines.

Table 10*Intercoder Agreement*

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
INTERACTIONAL FEATURES						
Recasts						
Recast move		8	5	13	61.54	-
Linguistic focus	All subcodes	16	0	16	100.00	1.00
Intensity of recast	All subcodes	16	0	16	100.00	1.00
Intensity of recast form	All subcodes	16	0	16	100.00	1.00
Completeness	All subcodes	16	0	16	100.00	1.00
Length	All subcodes	16	0	16	100.00	1.00
Uptake	All subcodes	14	3	17	82.35	0.79
Negotiation of meaning						
Trigger	All subcodes	64	2	66	96.97	0.94
Indicator		48	11	59	81.36	-
Initiator	All subcodes	66	0	66	100.00	1.00
Type	All subcodes	62	4	66	93.94	0.92
Response						
Type	All subcodes	60	6	66	90.91	0.90
Accuracy	All subcodes	36	0	36	100.00	1.00
Reaction to response	All subcodes	58	7	65	89.23	0.87
Complexity	All subcodes	42	0	42	100.00	1.00

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
Structural alignment						
Copula <i>be</i>		332	16	348	95.40	-
Direction of alignment	All subcodes	228	0	228	100.00	1.00
<i>Have</i> + noun phrase		110	8	118	93.22	-
Direction of alignment	All subcodes	76	2	78	97.44	0.97
<i>For</i> + noun phrase		214	2	216	99.07	-
Direction of alignment	All subcodes	128	4	132	96.97	0.96
Modal + infinitive		242	6	248	97.58	-
Direction of alignment	All subcodes	156	1	157	99.36	0.99
<i>To</i> -infinitive		94	0	94	100.00	-
Direction of alignment	All subcodes	32	0	32	100.00	1.00
Existential <i>there</i>		16	0	16	100.00	-
Direction of alignment	All subcodes	-	-	-	-	-
Question with subject-auxiliary inversion		104	11	115	90.43	-
Direction of alignment	All subcodes	70	0	70	100.00	1.00
<i>Wh</i> -interrogative		84	6	90	93.33	-
Direction of alignment	All subcodes	36	1	37	97.30	0.97
Lexical alignment						
Direction of alignment	All subcodes	104	3	107	97.20	0.97
COMPLEXITY AND ACCURACY						
AS-units	All subcodes	384	51	435	88.28	-
Requests		178	33	211	84.36	-

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
Request strategy	All subcodes	180	30	210	85.71	0.84
Internal modifier	All subcodes	38	3	41	92.68	0.85
External modifier	All subcodes	34	4	38	89.47	0.87
Appropriateness	All subcodes	184	26	210	87.62	0.81
COMMUNICATIVE ADEQUACY						
Task completion	All subcodes	16	0	16	100.00	1.00
Generic structure	All subcodes	14	2	16	87.50	0.81
Generic style	All subcodes	14	2	16	87.50	0.81

Note. For subcodes with a small number of occurrences in the double-coded sample, intercoder agreement is indicated for the overall agreement across all subcodes for a given category, to base the agreement measures on a reasonable number of codings. A complete table with agreement measures for each code can be found in the appendix (see Appendix I). For codes where the segments to be coded were defined beforehand (see multi-step coding procedure described above), the chance-corrected Brennan-Prediger kappa coefficient in addition to percentage agreement is also given. For codes that were not applied to pre-defined segments, only the percent agreement is indicated as chance agreement was unlikely.

Statistical Analyses

The data obtained from coding was subsequently statistically analyzed to answer the research questions. Since research question 1, focusing on the types and characteristics of interactional features, had a qualitative focus, the data relating to this research question was analyzed using descriptive statistics only: The coding counts of each interactional feature and related subcategories were counted and percentage frequencies were calculated. Since the two interaction types produced an unequal number of dialogs which were also of different lengths, the counts were also corrected for length by calculating ratios per message.

The data relating to research question 2, focusing on task performance differences between system and SCMC interaction, was analyzed using descriptive statistics (mean, standard deviation) as well as statistical tests to analyze differences in performance measures between the two interaction types. SPSS Statistics (Version 27) (*IBM SPSS Statistics, 2020*) was used to perform all statistical data analyses. Independent sample two-tailed *t*-tests were used on the interval-scaled measures. Preliminary Shapiro-Wilk tests were run to test if the data was normally distributed. Group differences for variables with data significantly deviating from a normal distribution were alternatively assessed using Mann-Whitney *U* tests. Variables with ordinal scales (e.g., rating scores) were analyzed using Mann-Whitney *U* tests.

Questionnaires

The pre- and post-questionnaire data was analyzed by first grouping the responses to the open-format questions and subsequently counting and calculating frequency counts of all responses.

Interviews

The learner interviews were transcribed verbatim by the author, including grammatical or lexical errors and filler words such as “äh” etc. Swiss German dialect was translated into Standard German, approximating language and punctuation to a written form of Standard German. Significant pauses were marked with bracketed periods, with the number of periods indicating the duration of the pause in seconds,

e.g., (.), (...). Longer pauses were indicated with the number of seconds in brackets, e.g., (5). Nonverbal activities and utterances were noted in double brackets, e.g., ((lacht)). Unintelligible speech was marked as (unv.). Abruptly interrupted/truncated utterances or words were marked by double hyphens, i.e., --.

The interview transcripts were subsequently analyzed by the author through content-structuring qualitative content analysis (Kuckartz, 2018) using MAXQDA 2020 (VERBI Software, 2020). A set of main codes (in English language) was initially established top-down based on aspects of the research questions, i.e., *negotiation of meaning, recasts, accuracy, complexity* etc. Relevant sequences in the interview transcripts were assigned to the main codes and subsequently refined into subcodes. The subcodes emerged bottom-up from the interview data.

Since the code development was closely linked to the interview guide in that the main codes were derived from the interview questions (e.g., in a text segment containing a question on task completion or negotiation of meaning, the student's response would be assigned the main code "task completion" or, respectively, "negotiation of meaning") and the subcodes were phrased closely to the original text by means of summarizing or paraphrasing, thus reducing the interpretative subjectivity when developing the code system, the interview data was not coded by a second coder to establish intercoder reliability.

The codes derived from the coding of the interviews will be presented in the chapter Findings, as they themselves are considered a finding of the study. As noted above, the analysis of the interviews did not include frequency counts as the goal was to capture the breadth of the learners' perceptions and reactions rather than to measure and compare the prevalence of a given phenomenon in the two groups (system, SCMC). So the findings on each code will not include frequency codes but instead specify if a given code occurred in the data of both groups (system, SCMC) or if it only occurred in one of the groups to highlight whether a given learners' perception was shared in both groups or whether it was specific to a given interaction type.

4 Findings

This chapter presents the findings from the data analysis described in the previous chapter. The findings will be organized in response to the research questions, which are restated below.

1. *What interactional features occur in written task-based interaction between commercial VET EFL learners and a dialog system and what are their characteristics compared to those found in learner-peer written SCMC on the same task?*
2. *Is there a difference in the learners' task performance in terms of linguistic complexity and accuracy as well as communicative adequacy between learner-system interaction and learner-peer SCMC?*

The chapter first presents a general overview of relevant key measures of the two interaction types (system, SCMC) side by side. This is followed by a section each on the findings related to research question 1 (interactional features) and the findings on research question 2 (task performance). Each section first presents the findings from the analysis of the interaction scripts, followed by the findings from the analysis of the supplementary data (questionnaires, interviews), and concluding with a summary of the key findings.

The findings will be interpreted and discussed in the chapter Discussion, also integrating and jointly interpreting the findings from the different data sources (scripts, questionnaires, interviews).

As a note on the sections relating to the questionnaire findings, the text on the questionnaire items including response options and responses will be presented in English (translated from the German original by the author) for better readability.

4.1 General Overview of Interactions

As a general overview of the two types of interaction (system and SCMC) Table 11 presents for both interaction types the following general measures, along with the results from independent samples two-tailed *t*-tests to check for statistical differences

between the two interaction types: mean time on task, mean total turns and total messages produced by both learner and interlocutor, and mean total messages and mean total words produced per learner. Since preliminary Shapiro-Wilk tests at level $\alpha = .05$ detected significant deviation from normality for the *words per learner* measure, the Mann-Whitney U test was used for this variable.

As noted above, *learner* consistently denotes the student in the role of hotel employee and *interlocutor* refers to the dialog system in learner-system interaction and the student peer in SCMC in the role of hotel employee.

As can be seen in Table 11, learners who interacted with the dialog system spent significantly less time (in min) on task ($M = 13.05$, $SD = 3.68$) than learners interacting with a peer through SCMC ($M = 16.78$, $SD = 3.14$), with a large effect size ($d = -1.09$).

Despite less time interacting, learner-system interaction elicited a significantly larger amount of text than SCMC: Total turns per interaction produced by both learner and interlocutor were significantly higher for system interaction ($M = 45.33$, $SD = 17.96$) than for SCMC ($M = 24.11$, $SD = 11.16$), corresponding to a large effect size ($d = 1.42$). Total messages per interaction produced by both learner and interlocutor were also significantly higher in learner-system interaction ($M = 54.24$, $SD = 20.86$) than in SCMC ($M = 33.94$, $SD = 15.99$). The effect size was also large ($d = 1.09$).

Considering only learner-produced messages, the number was also significantly higher in system interaction ($M = 22.67$, $SD = 8.98$) than in SCMC ($M = 15.61$, $SD = 8.12$), corresponding to a large effect size ($d = 0.83$). Similarly, the learner-produced number of words per interaction was significantly higher for learner-system interaction ($M = 161.86$, $SD = 50.86$) than for SCMC ($M = 119.44$, $SD = 30.65$), with a large effect size ($d = 1.01$).

Table 11

Descriptive Statistics and Results of Independent Sample T-Test (Two-tailed) and Mann-Whitney U Test for Differences in General Measures Between Learner-System Interaction and SCMC

Measure	M (SD)		t	Mann-Whitney U	p	Cohen's d
	System (n = 21)	SCMC (n = 18)				
Time on task in min	13.05 (3.68)	16.78 (3.14)	-3.38	-	.002**	-1.09
Total turns produced by learner and interlocutor	45.33 (17.96)	24.11 (11.16)	4.50	-	<.001***	1.42
Total messages produced by learner and interlocutor	54.24 (20.86)	33.94 (15.99)	3.36	-	.002*	1.09
Total messages produced by learner	22.67 (8.98)	15.61 (8.12)	2.56	-	.015*	0.83
Total words produced by learner ^b	161.86 (50.86)	119.44 (30.65)	-	91.50	.006*	1.01

^a Due to significant ($p < .05$) lack of variance homogeneity, results from the Welch's test are reported for this variable. ^b Due to non-normally distributed data, the Mann-Whitney U test was conducted for this variable.

* $p < .05$. ** $p < .01$. *** $p < .001$.

4.2 Research Question 1: Interactional Features

This section presents the findings with regard to research question 1 on the occurrence and characteristics of interactional features found in learner-system interaction compared to SCMC. The findings will be summarized at the end of the section.

As described in the previous chapter, coding the scripts revealed the following interactional features: negotiation of meaning, recasts, and lexical and structural alignment. As also described in the previous chapter, each type of interactional feature was further assigned subcodes relating to their specific characteristics. These subcodes were either taken from the existing literature or were newly developed based on the data. As argued, both the newly developed subcodes as well as those applied from the

literature on human-human interaction reveal the specific characteristics of learner-system interactional features and are therefore reported as a finding. Each subcode will be presented and illustrated with examples from the data. Differences in the realization of codes previously applied to human-human interaction will be highlighted and discussed in the chapter Discussion. The number of occurrences for each code and subcode will also be reported, allowing for quantitatively characterized patterns in the occurrence of interactional features.

Furthermore, the number and nature of coded segments found in learner-system interaction will be contrasted with those found in learner-peer SCMC on the same task. As argued above (see section Research Gaps and Derivation of Research Questions), contrasting the interactional features of the two interaction types aided in highlighting the features' characteristics unique to the learner-system interaction context. As described in the same section, this was done by applying the same codes that had been developed based on the learner-system interactions to the SCMC scripts. The counts and detailed description of the codes found in SCMC will be presented in comparison with those from the learner-system interactions at the end of each section dedicated to the different interactional features.

4.2.1 Negotiation of Meaning

Negotiation of Meaning in Learner-System Interaction

As outlined above (section Coding of Negotiation of Meaning), negotiation of meaning was defined as an episode where an explicit indication of non-understanding occurs, and the course of conversation is halted while the interlocutors attempt to resolve the communication breakdown.

Negotiation of meaning was found frequently in learner-system interaction. Table 12 shows the number of times non-understanding was indicated by the learner or interlocutor (i.e., dialog system) across the 21 learner-system interactions. As can be seen, a total of 218 indicators or 10.38 indicators per interaction and 0.19 indicators per message were identified. In other words, 19% of all learner- and system-produced messages consisted of an indication of non-understanding.

Table 12*Frequency of Indicators of Negotiation of Meaning in Learner-System Interaction*

Measure	Frequency
Total indicators	218
Indicators per interaction	10.38
Indicators per message	0.19

Note. Number of learner-system interactions $n = 21$, number of messages in learner-system interactions $n = 1139$.

As explained in the chapter Methods (see section Coding of Negotiation of Meaning), when coding negotiation of meaning episodes, each of their four components (i.e., trigger, indicator, response, reaction to response) was assigned subcodes relating to their specific characteristics. An overview of all subcodes relating to the negotiation components along with their counts for learner-system interaction and SCMC is presented in Table 13. The following paragraphs will present the subcodes for each component in detail below along with definitions and examples. Since the analysis of the indicator component introduces aspects that will be relevant to the analysis of the trigger component, the findings on the indicator will be presented first for better clarity. Definitions and coding guidelines with additional examples can also be found in the codebook (for link, see Appendix H).

Table 13
Overview of Main and Subcodes for Negotiation of Meaning with Counts

Main code	Subcode	Count	
		System (<i>n</i> = 21)	SCMC (<i>n</i> = 18)
Trigger			
Accuracy	Erroneous	121	6
	Target-like	97	3
Indicator			
Initiator	Learner	4	4
	Interlocutor	214	5
Type	Clarification request	101	1
	Confirmation check	8	2
	Inappropriate response	109	2
	Overlap	-	4
Response			
Type	Meta-comment	21	3
	Change topic	55	-
	Accept topic	49	3
	Elaboration	35	2
	Reduction	24	-
	Modification	23	-
	Repetition	4	-
	No response	7	-
Accuracy of reformulation	Improved	33	-
	Reduced	26	-
	Equivalent	23	2
Reaction to response	Implicit resolution	92	5
	Explicit resolution	-	4
	Implicit no resolution	56	-
	Explicit no resolution	63	-
	No reaction to response	7	-
Complexity	Single-layered	139	7
	Multi-layered	33	1

Indicator

As described in the chapter Literature Review (see section Negotiation of Meaning), the indicator describes the part of the negotiation episode where one of the interlocutors signals non-understanding.

Initiator

In the current analysis, the indicator component was assigned a subcode that reflected whether the initiator of the negotiation sequence was the learner or the

interlocutor. As shown in Table 14, almost all negotiation of meaning episodes in learner-system interaction were initiated by the interlocutor, i.e., the dialog system, and only four indicators were uttered by the learner.

Table 14
Frequency and Distribution of Subcodes for Initiator of Negotiation of Meaning in Learner-System Interaction

Subcode	Definition	Example	Count	%
Initiator				
Learner	The learner initiates the negotiation sequence, i.e., indicates non-understanding.	System: So, I will book room Alpha on the 12 october for you. Could you please confirm if you would like me to proceed? Learner: <u>What is proceed?</u>	4	1.8
Interlocutor	The interlocutor initiates the negotiation sequence, i.e., indicates non-understanding.	Learner: thats nice System: <u>Excuse me, can you please rephrase your last message?</u>	214	98.2

Type

Three indicator subcodes were identified in the learner-system data, representing the different types of indicators learners and interlocutor used to signal non-understanding. Table 15 provides an overview of the indicator subtypes along with definitions and examples. As pointed out in the Method chapter as well as in the overview in Table 14, this classification of indicator types was adopted from the literature on negotiation of meaning in human-human interaction. The following paragraphs illustrate each indicator type with examples from the study data. Differences in the realization of the indicator types found in learner-system interaction compared to that found in the literature on human-human interaction will be highlighted in the chapter Discussion.

The overview in Table 15 also shows the total and relative frequencies of each indicator type. As seen, the most frequent types of indicators were inappropriate responses (109 indicators, 50.0%) and clarification requests (101 indicators, 46.3%).

Confirmation checks were produced infrequently to indicate non-understanding (eight indicators, 3.7%).

With respect to the four learner-produced indicators, three of them were realized as clarification requests, while one was a confirmation check.

Table 15

Frequency and Distribution of Subcodes for Type of Indicator of Negotiation of Meaning in Learner-System Interaction

Subcode	Definition	Example	Count	%
Type				
Clarification request	A type of indicator that seeks to elicit clarification of the interlocutor's preceding utterance.	Learner: ohh thats nice System: <u>Excuse me, can you please rephrase your last message?</u>	101	46.3
Confirmation check	A type of indicator that serves to establish that the previous utterance was understood correctly.	Learner: Yes I confirm to book room Alpha for the 12 octobre 2019 System: <u>Please could you confirm if you would like to book room Alpha for the 12 octobre 2019?</u>	8	3.7
Inappropriate response	The indicator utterance signals indirectly and implicitly that something was not understood by giving an inappropriate response to the previous utterance.	Learner: Has the Beta room also got monitors and speakers? System: <u>The Beta room has space for a total of 190 people, indoor and outdoor.</u>	109	50.0

Analyzing the realizations of each indicator type in more detail revealed the following additional observations: In most clarification requests, the interlocutor asked the learner to “rephrase” or “say” or “try” again the preceding message, as illustrated in Example 14. So, clarification requests generally prompted the learner to clarify the entire content of their preceding utterance rather than for example a specific lexical item of their utterance.

Example 14*Clarification request entire content*

- | | | | |
|----|---------|---|------------------------------|
| 79 | Learner | How expensiv is the Room alpha? | |
| 80 | System | Excuse me, can you please rephrase your last message? | <i>Clarification request</i> |

Only in a few instances (14 instances), the system indicated partial non-understanding requesting the learner to specify the room name (Example 15).

Example 15*Clarification request partial non-understanding*

- | | | | |
|----|---------|---|------------------------------|
| 11 | Learner | I need to make a presentation | |
| 12 | System | Could you tell me which room you are interested in? | <i>Clarification request</i> |

In technological terms, system-produced clarification requests on the entire content of the preceding message refer to the system's *fallback messages* which were issued in one of the following situations (see section Dialog System

One, when the learner message could not be assigned to an intent with sufficiently high confidence. Two, due to error detection which became effective when the spelling error-to-word ratio of a learner message reached or exceeded 25%.

In two clarification requests, the dialog system asked the learner to restate a specific piece of information from the preceding utterance. These instances occurred when the system initially failed to extract an entity value in the learner's message such as a name (Example 16) or date (Example 17).

Example 16*Clarification request for specific piece of information*

- | | | |
|---|---------|--|
| 4 | Learner | Hello i'm Dariana from the company of a marketing. I have question. I need a room for 150 people for a presentation with a great area. And also i want more Information of you for a good room for our presentation. |
|---|---------|--|

- | | | | |
|---|--------|--|------------------------------|
| 5 | System | Thank you for contacting One King Hotel. I will gladly assist you in finding a room. | |
| 6 | System | Please could you give me your name? | <i>Clarification request</i> |

Example 17*Clarification request for specific piece of information*

- | | | | |
|---|---------|---|------------------------------|
| 1 | Learner | Greetings from Swizerland, we want to book a meeting Room the 12 Oktober in your hotel is it possible ? | |
| 2 | System | Could you please give me the date for the reservation? | <i>Clarification request</i> |

As stated above, three of the total four learner-produced indicators were of the clarification request type. In one instance the learner requested clarification on the entire content of the preceding message, as illustrated in Example 18.

Example 18*Clarification request by learner*

- | | | | |
|----|---------|--|------------------------------|
| 13 | System | The Beta room has space for a total of 190 people, indoor and outdoor. | |
| 14 | Learner | Hä | <i>Clarification request</i> |

In the other two instances, the learners requested clarification about a specific lexical item. This is illustrated in Example 19, where the learner asks in what way the interlocutor would like them to “confirm” their reservation.

Example 19*Clarification request by learner for specific piece of information*

- | | | | |
|----|---------|--|------------------------------|
| 51 | System | So, I will book room Gamma on the 13 of july 2020 for you. Could you please confirm if you would like me to proceed? | |
| 52 | Learner | per mail or like what | <i>Clarification request</i> |

Confirmation checks serve to establish that the preceding utterance was understood correctly. The confirmation checks that were produced by the dialog system exclusively sought to confirm the booking date the learner had stated in the preceding utterance, as illustrated in Example 20.

Example 20

Confirmation check

- | | | | |
|----|---------|--|---------------------------|
| 52 | Learner | I take room Alpha for the 12th october 2019. | |
| 53 | System | So, I will book room Alpha on the 12th october 2019 for you. Could you please confirm if you would like me to proceed? | <i>Confirmation check</i> |

The only learner-produced confirmation check involved the learner's testing of a hypothesis on the interlocutor's problematic utterance. The extract is presented in Example 21. As can be seen, the learner deduced from the system's messages following their own reservation request that it might require information to complete the reservation. The learner consequently tests their hypothesis by asking if the system needs more information.

Example 21

Confirmation check by learner

- | | | | |
|----|---------|--|---------------------------|
| 38 | Learner | The reservation is for 12 october 2019 | |
| 39 | System | Please could you confirm if you would like to book room Alpha for the 12th october 2019? | |
| 40 | System | Room Alpha has a big screen and a state-of-the-art projector and audio system. | |
| 41 | System | It is possible to add a portable screen as well as projector and loudspeakers to Beta. | |
| 42 | Learner | Do you need more informations? | <i>Confirmation check</i> |

Inappropriate responses, i.e., indicators signaling non-understanding implicitly by giving an inappropriate response to the previous utterance, that were produced by

the dialog system were, in technological terms, caused by the system's incorrect intent classification. The misclassification of intents resulted in responses with varying degrees of inappropriateness. Inappropriate responses that involved the dialog system correctly extracting the entity, for example, the room name or topic of a message while misclassifying its intent, may seem less severe. This led to inappropriate responses on the correct topic (Example 23) or room (Example 24), which might seem less inappropriate than instances without entity extraction (Example 22).

Example 22

Inappropriate response

77	System	What would be the date for the booking?	
78	Learner	12 oktober	
79	System	Room Alpha has space for up to 270 people.	<i>Inappropriate response</i>

Example 23

Inappropriate response with correct entity extraction

21	Learner	no, which room is the best for have also place for tidy conversation	
22	System	All three rooms are supplied with the necessary equipment to hold a presentation.	<i>Inappropriate response</i>

Example 24

Inappropriate response with correct entity extraction

12	Learner	Has the Beta room also got monitors and speakers?	
13	System	The Beta room has space for a total of 190 people, indoor and outdoor.	<i>Inappropriate response</i>

Trigger

The trigger component in a negotiation of meaning episode refers to the utterance or part of an utterance that creates non-understanding (see section Negotiation of Meaning in the chapter Literature Review). Previous studies on human-

human interaction have classified triggers according to the element or aspect that caused the non-understanding, such as a specific lexical item, grammatical structure, or unclear reference of a pronoun (Smith, 2003). This is possible when the specific problematic element is identified in the indicator as in, e.g., “What does *janitor* mean?”

In the learner-system interaction data of the present study, however, the indicator in most cases did not identify which element of the trigger utterance caused non-understanding, as will be shown in more detail below. For instance, the indicator type *inappropriate response*, being unrelated to the preceding message, naturally contained no indication on the element to which the non-understanding could be attributed. Similarly, the *clarification request* type issued by the dialog system did not allow to determine which element of the learner’s message was problematic as it referred to the entire learner message with the pronoun “that” (see Example 25) or by explicitly stating “your last message” (see Example 26). The added request to rephrase is equally unspecific, prompting the learner to “try one more time” or to “say it again” (see Example 25 and Example 26).

Example 25

Clarification request without specifying the problematic element in the trigger

12	Learner	There are some more critereas i would like to say	<i>Trigger</i>
13	System	I am sorry, I did not understand that. Could you please try one more time?	<i>Clarification request</i>

Example 26

Clarification request without specifying the problematic element in the trigger

26	Learner	Okay and how much does it costs?	<i>Trigger</i>
27	System	So sorry, I could not understand your last message. Could you please say it again?	<i>Clarification request</i>

Only a small number of clarification requests, as explained above in the section on indicators, pointed to the specific content element that needed clarification, such as the room name or learner’s name. The examples from the previous section for this

group of clarification requests are presented again below. As can be seen in these examples, the element of the content that needs clarification (i.e., room name, customer name, booking date) is clearly stated. However, it remains unclear which element of the learner's utterance (e.g., lexical item, syntactical structure, discourse structure) caused the problem in the dialog system's understanding of the element.

Example 27

Clarification request specifying problematic element in the trigger

11	Learner	I need to make a presentation	<i>Trigger</i>
12	System	Could you tell me which room you are interested in?	<i>Clarification request</i>

Example 28

Clarification request specifying the problematic element in the trigger

4	Learner	Hello i'm Dariana from the company of a marketing. I have question. I need a room for 150 people for a presentation with a great area. And also i want more Information of you for a good room for our presentation.	<i>Trigger</i>
5	System	Thank you for contacting One King Hotel. I will gladly assist you in finding a room.	
6	System	Please could you give me your name?	<i>Clarification request</i>

Example 29

Clarification request specifying the problematic element in the trigger

1	Learner	Greatings from Swizerland, we want to book a meeting Room the 12 Oktober in your hotel is it possible ?	<i>Trigger</i>
2	System	Could you please give me the date for the reservation?	<i>Clarification request</i>

Similarly, the indicator type *confirmation check* explicitly highlighted the booking date as the part of the utterance that was not understood correctly but failed to indicate which element of the learner's message had caused the non-understanding. This becomes evident in Example 30: While the dialog system's confirmation check makes it clear that the learner needs to confirm the booking for the given date, it does not specify which element of the learner's message made the reconfirmation necessary.

Example 30

Confirmation check specifying the problematic element in the trigger

52	Learner	I take room Alpha for the 12th october 2019.	<i>Trigger</i>
53	System	So, I will book room Alpha on the 12th october 2019 for you. Could you please confirm if you would like me to proceed?	<i>Confirmation check</i>

So, the cause of the non-understanding could generally not be attributed to a specific lexical item or grammatical construction in the trigger message. Following Smith's (2003) categorization of negotiation triggers, all triggers found in the learner-system data of this study could thus be classified as *content trigger* type, i.e., "instances where the entire content of a previous message is in some way problematic" or "the problem could not be attributed to one of the other trigger types" (p. 43).

The only exception was found in a learner-initiated negotiation episode shown in Example 31. As seen, in this extract from a learner-initiated negotiation episode where the learner identified a specific lexical item as cause for their clarification request:

Example 31

Clarification request by learner specifying the problematic item in the trigger

43	System	So, I will book room Alpha on the 12 october for you. Could you please confirm if you would like me to proceed?	<i>Trigger</i>
44	Learner	What is proceed?	<i>Confirmation check</i>

Accuracy

Instead, triggers were classified according to their accuracy by assigning the subcodes *target-like* or *erroneous*. This was considered an important distinction because target-like learner utterances that provoke non-understanding in the interlocutor might be perceived as ‘false positives’ by the learner. On the other hand, erroneous utterances that cause no non-understanding function as ‘false negatives’, constituting missed opportunities for negative feedback.

As shown in Table 16, target-like learner utterances were almost equally likely to function as triggers as erroneous utterances.

Table 16
Frequency and Distribution of Subcodes for Accuracy of Trigger of Negotiation of Meaning in Learner-System Interaction

Subcode	Definition	Example	Count	%
Accuracy				
Erroneous	The trigger, i.e., utterance in a negotiation of meaning episode that creates non-understanding, contains at least one grammatical, lexical, or spelling error and/or discourse error (i.e., unclear meaning or reference, missing contextual relevance).	Learner: <u>Which room gives a good atmosphere?</u> System: I am afraid I have trouble understanding. Please could you rephrase?	121	55.5
Target-like	The utterance that creates non-understanding is target-like.	Learner: <u>i like all 3 rooms. can you help me to decide?</u> System: So sorry, I could not understand your last message. Could you please say it again?	97	44.5

As stated in the definition (Table 16), utterances were coded as *erroneous* if they contained at least one linguistic, i.e., grammatical, lexical, or spelling error (Example 32, Example 33, Example 34) or discourse error, i.e., unclear meaning or reference, missing contextual relevance (Example 35, Example 36). As seen in the examples, discourse errors typically involved the learner ignoring the dialog system’s

preceding utterance which typically occurred when the dialog system's utterance itself lacked reference or relevance with regard to the learner's preceding message. The classification of linguistic errors followed the operationalizations outlined above (section Coding of Errors). It should be noted that the category included all errors regardless of their severity, i.e., utterances with minor errors such as the incorrect inflection of "cost" in Example 34 as well as utterances with more severe errors such as the misselection of the word "speaks" instead of "conversation", as in Example 32.

Example 32

Erroneous trigger with lexis and grammar errors

17	Learner	In which room can i <u>make also</u> [lexis error; grammar error] private <u>speaks</u> [lexis error]	<i>Trigger</i>
18	System	I am sorry, I did not understand that. Could you please try one more time?	<i>Indicator</i>

Example 33

Erroneous trigger with spelling and grammar errors

14	Learner	Which room is the best for <u>smalltalks</u> [spelling error; grammar error] and <u>have</u> [grammar error] a good <u>atmosphere</u> [spelling error]?	<i>Trigger</i>
15	System	Yes, absolutely.	
16	System	Could you please tell me which room you are referring to?	<i>Indicator</i>

Example 34

Erroneous trigger with grammar error

26	Learner	Okay and how much does it <u>costs</u> [grammar error]?	<i>Trigger</i>
27	System	Yes, absolutely.	<i>Indicator</i>

Example 35

Erroneous trigger with discourse error

25	Learner	yes i would like to know how the hotel is	
----	---------	---	--

26	System	I would recommend that you take the subway to our hotel.	
27	Learner	and it costs 1100?	<i>Trigger</i>
28	System	I am afraid I have trouble understanding. Please could you rephrase?	<i>Indicator</i>

Example 36

Erroneous trigger with discourse error

18	Learner	The 12th October 2019	
19	System	I will reserve room None for the 12th October 2019. Could you please confirm if this is correct?	
39	Learner	Is this room comfortable?	<i>Trigger</i>
40	System	Beta is CHF 900 including the terrace.	<i>Indicator</i>

Target-like utterances, on the other hand, contained no linguistic or discourse errors as illustrated in Example 37.

Example 37

Target-like trigger

15	System	So, I will book room Alpha on the 12th October 2019 for you. Could you please confirm if you would like me to proceed?	
16	Learner	I confirm the booking of room Alpha, thanks	<i>Trigger</i>
17	System	Room Alpha is equipped with a big screen, a sound and lighting system as well as a stage.	<i>Indicator</i>

Technologically, there were two reasons for the inconsistent classification of erroneous learner utterances as triggers. As explained above (see section Dialog System), the system produced an indicator of non-understanding based on three mechanisms: 1) error detection, 2) incorrect intent assignment, 3) confirmation of a booking date. All three mechanisms only partly depended on the linguistic accuracy of the utterance: With regard to error detection, the mechanism only considered spelling errors and, in addition, only became effective if the error-to-word ratio met or

exceeded 25%. As a consequence, an erroneous utterance would not systematically lead to an indicator, since utterances whose relative error frequency was below the threshold or contained errors other than spelling errors might not have set off error detection. With respect to the other two sources of indications of non-understanding, intent assignment and date confirmation both depended on the linguistic accuracy of an utterance only to a certain degree, in that the dialog system was trained on linguistically accurate data, i.e., was trained to ‘understand’ linguistically accurate language. In other words, only utterances that deviated too far from the modeled language might have caused an issue to the system’s intent classification. However, as the NLU’s intent classification additionally relied on other factors such as context, perfectly target-like utterances may have been incorrectly classified by the system.

Response

The response is the negotiation component following the indicator and acknowledging the indication of non-understanding (see section Negotiation of Meaning in the chapter Literature Review).

As will be described in more detail in the following paragraphs, in the present learner-system interaction data, not all indicators were followed by a response in the sense of the above definition. Instead, the indicator may be followed by an utterance coded as ‘*non-response*’, i.e., an utterance that served to exit rather than complete the negotiation routine via one of the strategies that will be described further below: *meta-comment*, *topic abandonment*, *no response*. To identify and highlight the discrepancies between the response types previously documented in human-human interaction and those observed in the system data of this study, these ‘non-responses’ were also analyzed and discussed under the term *response*.

The visualization in Figure 14 shows the different paths the learner and dialog system may follow after being confronted with an indicator of non-understanding. As shown, the indicator may be followed by a response in the sense of the definition above, i.e., acknowledging and responding to the non-understanding highlighted in the indicator. Alternatively, it may be followed by a ‘non-response’ such as a meta-comment, topic change or acceptance, or no response at all.

Figure 14

Visualization of Possible Paths Following an Indicator of Non-understanding in Negotiation of Meaning in Learner-System Interaction

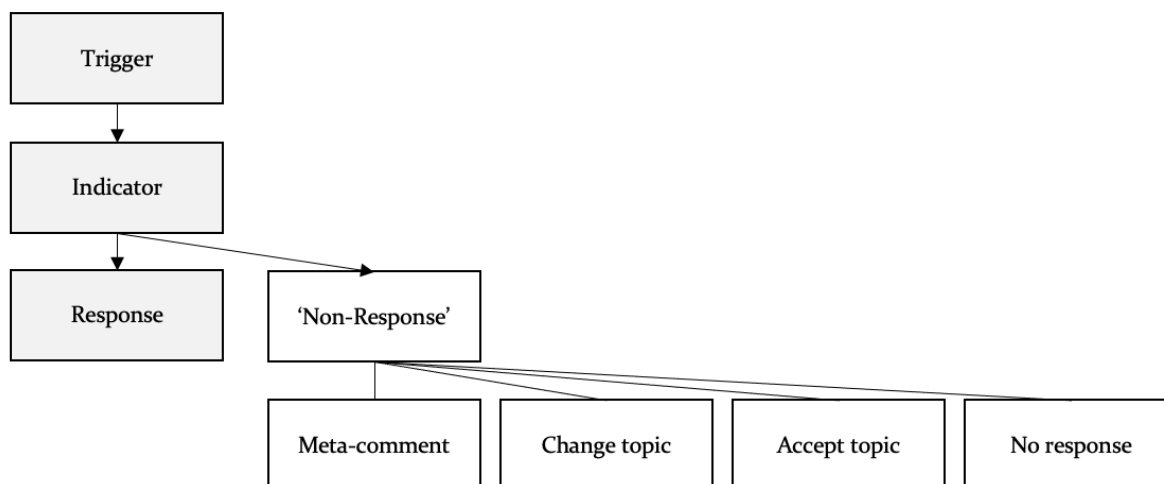


Table 17 lists the response types including ‘non-responses’ found in the learner-system data along with definitions and examples as well as counts and percentage distribution indicating the prevalence of each type. Each response type will be described in more detail below.

Table 17

Frequency and Distribution of Subcodes for Type of Response of Negotiation of Meaning in Learner-System Interaction

Subcode	Definition	Example	Count	%
'Non-responses'			132	60.6
Meta-comment	The respondent responds to the indicator on a metalinguistic level commenting either on the indicator or overall non-understanding,	Learner: thank you very much, have a nice day [trigger] System: I am sorry, I did not understand that. Could you please try one more time? [indicator] Learner: <u>Nevermind</u> [response]	21	9.6
Change topic	The respondent abandons the topic of the message that created non-understanding	Learner: How many People can stay here? [trigger] System: I am afraid I have trouble understanding.	55	25.2

Subcode	Definition	Example	Count	%
	and continues with a new topic.	Please could you rephrase? [indicator] Learner: <u>is this the One King Hotel?</u> [response]		
Accept topic	The respondent abandons the topic of the message that created non-understanding and continues with the topic that the interlocutor introduced during the indicator move.	Learner: First name? [trigger] System: How many guests are you expecting? [indicator] Learner: <u>About 150 or more</u> [response]	49	22.5
No response	There is no response given at the end of an interaction or due to message overlap.	Learner: You booked for me so already the room Alpha. [trigger] System: Excuse me, could you rephrase this please? [indicator] [No response/end of interaction]	7	3.2
Responses			86	39.4
Elaboration	The response provides more specific information by means of elaboration.	Learner: Hello, is this the One King Hotel in Toronto? [trigger] System: Excuse me, could you rephrase this please? [indicator] Learner: <u>I asked, if this is the "One King Hotel" in Toronto.</u> [response]	35	16.1
Reduction	The response reduces the information content of the utterance that caused non-understanding.	Learner: That sounds great. Is the room gamma fancy or like a normal room [trigger] System: Very well. Would you like to book room Gamma? [indicator] Learner: <u>Is room gamma too fancy?</u> [response]	24	11.0
Modification	The response involves formal modification (e.g., error correction, rephrasal by use of synonyms) of the utterance that caused non-understanding without expanding or reducing its information content.	Learner: I need a room for a event with my company. [trigger] System: I am sorry, I did not understand that. Could you please try one more time? [indicator]	23	10.6

Subcode	Definition	Example	Count	%
		Learner: <u>I need a room for an event with my colleagues.</u> [response]		
Repetition	The response involves exact repetition of the utterance that caused non-understanding.	Learner: and the others? [trigger] System: Room Gamma has two moveable monitors and speakers.[indicator] Learner: <u>and the others?</u> [response]	4	1.8

As shown in Table 17, of the overall 218 indicators, 132 or 60.6% were followed by a ‘non-response’ that exited a negotiation sequence. Responses that continued the negotiation sequence occurred in 86 or 39.4% of instances.

In meta-comments, a first type of ‘non-response’ occurring 21 times (9.6%), the respondent responded to the indicator on a metalinguistic level, commenting on the indicator or the overall non-understanding, as illustrated in the examples below. As the examples show, in meta-comments, the respondent typically expressed frustration, annoyance or bemusement over the system’s lack of understanding, as illustrated in Example 38 and Example 39.

Example 38

‘Non-response’ type Meta-comment

80	Learner	How expensiv is the Room alpha?	<i>Trigger</i>
81	System	Excuse me, can you please rephrase your last message?	<i>Indicator</i>
82	Learner	i think we have a probleme	<i>Meta-comment</i>

Example 39

‘Non-response’ type Meta-comment

56	Learner	and what about having apero?	<i>Trigger 1</i>
57	System	Room Gamma still has the original decoration from the 1920s. It is located on	<i>Indicator 1</i>

		the first floor of the hotel and offers plenty of space.	
58	Learner	thats not what i asked	<i>Meta-comment 1 / Trigger 2</i>
59	System	Thank you very much. Goodbye.	<i>Indicator 2</i>
60	Learner	wow, very polite	<i>Meta-comment 2</i>

In a few rare instances, respondents appeared to use meta-comments in an attempt to resolve non-understanding, as seen in Example 40.

Example 40

'Non-response' type Meta-comment as attempt to resolve non-understanding

38	Learner	The reservation is for 12 october 2019	<i>Trigger</i>
39	System	Please could you confirm if you would like to book room Alpha for the 12 october 2019?	
40	System	Room Alpha has a big screen and a state-of-the-art projector and audio system.	
41	System	It is possible to add a portable screen as well as projector and loudspeakers to Beta.	<i>Indicator</i>
42	Learner	Do you need more informations?	<i>Meta-comment</i>

In other instances, the topic of the utterance that had created non-understanding was abandoned by the respondent either by changing the topic of the conversation (Example 41) or by accepting the topic that the interlocutor had introduced in the indicator (Example 42).

Example 41

'Non-response' type Change topic

47	System	Room Beta can be equipped with a moveable screen, projector, and loudspeakers.	
48	Learner	Amazing	<i>Trigger</i>
49	System	Excuse me, can you please rephrase your last message?	<i>Indicator</i>

50	Learner	how much does the room beta coast?	<i>Change topic</i>
----	---------	------------------------------------	---------------------

Example 42*'Non-response' type* Accept topic

14	Learner	There are some more points	<i>Trigger</i>
15	System	Please could you tell me who I am chatting with?	<i>Indicator</i>
16	Learner	This is Mithusha	<i>Accept topic</i>

No response describes instances where the respondent ended the interaction after the interlocutor indicated non-understanding. This response type typically occurred in instances when the task had been formally completed (i.e., a room had been booked) with the learner's final utterance triggering an indicator of non-understanding which the learner ignored, as illustrated in Example 43.

Example 43*'Non-response' type* No response

36	Learner	No thats all.	<i>Trigger</i>
37	System	The cost for Alpha would be CHF 1400.	
38	System	Room Beta comes at a cost of CHF 900. This includes the price for the terrace.	
39	System	Room Gamma would be CHF 1100. [end of interaction]	<i>Indicator</i> <i>No response</i>

The responses as per the definition previously used in the literature that functioned to continue the negotiation sequence were further categorized in terms of their type. These corresponded to the strategy used to linguistically transform the trigger message that had caused non-understanding.

As can be seen in Table 17, the most frequent response type was *elaboration*, followed by *reduction* and *modification*. *Repetitions* were relatively infrequent. The response type *elaboration* refers to instances where the learner provided more specific

information to their original trigger utterance. This may be done by adding further information to the original utterances, as illustrated in Example 44 and Example 45.

Example 44

Response type Elaboration by adding further information

53	Learner	I confirm	<i>Trigger</i>
54	System	Excuse me, can you please rephrase your last message?	<i>Indicator</i>
55	Learner	Yes i confirm it i would like you to proceed	<i>Elaboration</i>

Example 45

Response type Elaboration by adding further information

19	Learner	Does room Beta have a projector ?	<i>Trigger</i>
20	System	So sorry, I could not understand your last message. Could you please say it again?	<i>Indicator</i>
21	Learner	I was asking if the room Betta have a Projecter, we won't to show are Gest a Video ?	<i>Elaboration</i>

The learner may also elaborate by detailing the attributes of the referent of the trigger utterance, as shown in Example 46.

Example 46

Response type Elaboration by detailing the attributes of the referent

69	Learner	I search a room	<i>Trigger</i>
70	System	I am afraid I have trouble understanding. Please could you rephrase?	<i>Indicator</i>
71	Learner	i search a room with place for 150 people	<i>Elaboration</i>

In other instances, the learner's elaboration involved specifying the room name – either by replacing the pronoun with the room name (Example 47) or by separately stating the room name (Example 48).

Example 47*Response type Elaboration by specifying the room name*

26	Learner	Okay and how much does it costs?	<i>Trigger</i>
27	System	So sorry, I could not understand your last message. Could you please say it again?	<i>Indicator</i>
28	Learner	How much does the Alpha room costs?	<i>Elaboration</i>

Example 48*Response type Elaboration by specifying the room name*

29	Learner	I want a room with the possibility of making presentations and having nice time with my collueges. it also need space for 150 person	<i>Trigger</i>
30	System	Please could you give me the name of the room?	<i>Indicator</i>
31	Learner	Gamma	<i>Elaboration</i>

Elaboration may also involve restating the original utterance more narrowly by using more specific vocabulary (Example 49 and Example 50):

Example 49*Response type Elaboration by using more specific vocabulary*

24	Learner	What are the qualities of the Gamma room?	<i>Trigger</i>
25	System	Room Gamma would be CHF 1100.	<i>Indicator</i>
26	Learner	And what about the atmosphere in room Gamma?	<i>Elaboration</i>

Example 50*Response type Elaboration by using more specific vocabulary*

6	Learner	Yes i already looked up what rooms you've got.	<i>Trigger</i>
7	System	Excuse me, could you rephrase this please?	<i>Indicator</i>
8	Learner	I have visited your web page and saw the rooms.	<i>Elaboration</i>

Finally, *elaboration* may involve a combination of the above strategies, e.g., specifying the room type along with the use of more specific vocabulary, as illustrated in Example 51.

Example 51

Response type Elaboration by combining strategies

6	Learner	I need to make a presentation	<i>Trigger</i>
7	System	Could you tell me which room you are interested in?	<i>Indicator</i>
8	Learner	Both. Do they have a beamer in?	<i>Elaboration</i>

Reduction refers to a response type where the learner reduced the information content of the trigger utterance, as illustrated in the examples below. *Reduction* may consist in repeating an unchanged fragment of the original trigger utterance, as seen in Example 52 and Example 53.

Example 52

Response type Reduction by reusing unchanged fragments from the trigger

19	Learner	We are 150 - 200 person and we come on the 12.10.2019	<i>Trigger</i>
20	System	So sorry, I could not understand your last message. Could you please say it again?	<i>Indicator</i>
21	Learner	we are 150	<i>Reduction</i>

Example 53

Response type Reduction by reusing unchanged fragments from the trigger

38	Learner	the date is at the 12. Oktober	<i>Trigger</i>
39	System	What would be the date for the booking?	<i>Indicator</i>
40	Learner	12. oktober	<i>Reduction</i>

More frequently, however, the reduced utterance reduction involved simultaneous modification of the original utterance, such as error correction, the replacing of lexical items, or syntactical restructuring. Example 54, Example 55, Example 56, and Example 57 illustrate reduction with modification.

Example 54

Response type Reduction with modification

5	Learner	Yeah what type of Meeting rooms do you hae ?v	<i>Trigger</i>
6	System	Could you please tell me which room you are referring to?	<i>Indicator</i>
7	Learner	What type of room do you have ?	<i>Reduction</i>

Example 55

Response type Reduction with modification

14	Learner	Which room is the best for smalltalks and have a good atmospher?	<i>Trigger</i>
15	System	Yes, absolutely.	
16	System	Could you please tell me which room you are referring to?	<i>Indicator</i>
17	Learner	Which room have the best atmosphere?	<i>Reduction</i>

Example 56

Response type Reduction with modification

30	Learner	we would like to come at the 12 june and then staying 5 days there	<i>Trigger</i>
31	System	Could you please give me the date for the reservation?	<i>Indicator</i>
32	Learner	the date for the reservation is the 12 june	<i>Reduction</i>

Example 57

Response type Reduction with modification

8	Learner	The should have space for at least 150 peoples, have a projector for a presentation, alow conversations insmall groups, have a good atmosphere, should	<i>Trigger</i>
---	---------	--	----------------

		look quite noble but not over the top and shouldn't cost more than 1500 Francs.	
9	System	So sorry, I could not understand your last message. Could you please say it again?	<i>Indicator</i>
10	Learner	Have you got any rooms with space for at least 150 Peoples?	<i>Reduction</i>

The final response type, *modification*, involved formal modification of the utterance that caused non-understanding without expanding or reducing its information content. The examples below illustrate this response type. *Modification* may involve a single transformation such as error correction (Example 58) or the replacement of a lexical item by a synonym (Example 59).

Example 58

Response type Modification with a single transformation

53	Learner	Goodbe	<i>Trigger</i>
54	System	I am sorry, I did not understand that. Could you please try one more time?	<i>Indicator</i>
55	Learner	goodbye	<i>Modification</i>

Example 59

Response type Modification with a single transformation

53	Learner	and does gamma have a beamer?	<i>Trigger</i>
54	System	The price for Gamma is 1100 CHF.	<i>Indicator</i>
55	Learner	does the room have a projector	<i>Modification</i>

Most modifications, however, involved a combination of transformations, such as spelling error correction combined with syntactical restructuring (Example 60) or the correction of a grammatical error along with synonym replacement (Example 61).

Example 60*Response type* Modification with a combination of transformations

32	Learner	how mutch for room alpha?	<i>Trigger</i>
33	System	Room Alpha is an auditorium with fixed seating and a stage.	<i>Indicator</i>
34	Learner	How much is Room Alpha?	<i>Modification</i>

Example 61*Response type* Modification with a combination of transformations

11	Learner	I would like to book a room for 1500 peoples.	<i>Trigger</i>
12	System	Excuse me, can you please rephrase your last message?	<i>Indicator</i>
13	Learner	I would like to reserve a room for 1500 people please.	<i>Modification</i>

Often, *modification* resulted in substantial or complete rephrasing of the original trigger utterance, as in Example 62 and Example 63.

Example 62*Response type* Modification with substantial rephrasing

17	Learner	In which room can i make also private speaks	<i>Trigger</i>
18	System	I am sorry, I did not understand that. Could you please try one more time?	<i>Indicator</i>
19	Learner	which room has the most place to have conversations	<i>Modification</i>

Example 63*Response type* Modification with complete rephrasing

23	Learner	But i will book a room, wich room do you think is the best for my options?	<i>Trigger</i>
24	System	Yes, sure.	

25	System	Could you please tell me which room you are referring to?	<i>Indicator</i>
26	Learner	i like all 3 rooms, can you help me to decide?	<i>Modification</i>

Accuracy of Reformulation

As an additional analysis, the accuracy of the response types involving reformulation (i.e., elaboration, reduction, modification) was assessed to gain an indication of learning behavior in students. This was done by assigning subcodes to the responses of the types *elaboration*, *reduction*, and *modification*: Whenever the response utterance contained fewer spelling, grammatical, lexical, or discourse errors than the original trigger utterance, its accuracy would be coded as *improved*. If the number of errors remained unchanged, either by retaining the original errors or correcting the original but adding new ones, the accuracy would be coded as *equivalent*. If the response utterance contained more errors, its accuracy would be coded as *reduced*.

Table 18 shows if the linguistic transformations through the means of elaboration, reduction, and modification resulted in utterances which were more or less accurate compared to the trigger utterance. As can be seen in the table, out of total 82 transformations, 33 or 40.2% led to improved accuracy. In 26 instances or 31.7%, the utterance remained equally accurate or inaccurate, respectively. About the same number of utterances, i.e., 23 or 28.0%, were less accurate than the trigger utterance.

Table 18

Frequency and Distribution of Subcodes for Accuracy of Response in Negotiation of Meaning in Learner-System Interaction

Subcode	Definition	Example	Count	%
Accuracy				
Improved	The response contains fewer (grammar, spelling, lexis, discourse) errors than the trigger utterance.	Learner: In which room can i make also private speaks [trigger] System: I am sorry, I did not understand that. Could you please try one more time? [indicator]	33	40.2

Subcode	Definition	Example	Count	%
Equivalent	The response contains more errors (i.e., grammatical, spelling, vocabulary, discourse errors) than the trigger utterance.	Learner: <u>which room has the most place to have conversations</u> [response]	26	31.7
		System: Would you like me to book room None for the None? [indicator]		
Reduced	The response contains the same number of errors (i.e., grammatical, spelling, vocabulary, discourse errors) as the trigger utterance.	Learner: <u>no, which room is the best for have also place for tidy conversation</u> [response]	23	28.0
		Learner: I confirm [trigger] System: Excuse me, can you please rephrase your last message? [indicator] Learner: <u>Yes i confirm it i would like you to proceed</u> [response]		

Note. Only the response types involving linguistic transformation of the trigger utterance were considered, i.e., elaboration, reduction, and modification, (n = 82).

Resolved vs. Non-Resolved by Response Type

As will be described in the next paragraph (see section Reaction to Response), the reaction to response component was assigned subcodes that indicated whether the non-understanding of the trigger utterance had been resolved (*resolution*) or not (*no resolution*). This allowed for a comparison of the success of the different response types in resolving non-understanding, which was considered a relevant analysis as it could be expected that a higher success rate of a specific response type may bring learners to gravitate to use them more frequently during system interaction over time.

Table 19 shows for each response type the frequency with which they led to a resolution vs. no resolution of the non-understanding from the trigger utterance. The overview reveals that elaborations were most successful in resolving non-understanding with 20 or 57.1% of total 35 elaboration responses being successful. Reductions and modifications were about equally successful, with 10 or 41.7% of 24 reductions and nine or 39.1% of 23 modifications, respectively, leading to a resolution of the non-understanding. Repetitions were the least conducive to resolving non-

understanding, with one or 25.0% of four repetitions leading to a resolution. Overall, slightly less than half of the responses (40 or 46.5%) resolved the non-understanding from the trigger utterance, whereas the other half did not resolve it (46 or 53.5%).

Table 19

Frequency and Distribution of Subcodes for Type of Response Leading to a Resolution vs. No Resolution of the Non-Understanding in Learner-System Interaction

Subcode	Resolution		No resolution	
	Count	%	Count	%
Elaboration	20	57.1	15	42.9
Reduction	10	41.7	14	58.3
Modification	9	39.1	14	60.9
Repetition	1	25.0	3	75.0
Total	40	46.5	46	53.5

Reaction to Response

The literature on human-human interaction has documented that negotiation routines may conclude with an optional reaction to response which describes the explicit comment by a speaker signaling understanding such as “okay” or “I get it now”. As will be described in more detail below, in the learner-system data of the present study, no negotiation component per this definition was observed. However, to further analyze how the negotiation routines were instead completed in the interaction data of this study, all utterances following the response move were coded as *reaction to response* moves and subsequently classified further into subcodes. As with the analysis of the response component (see above), this allowed for the identification and highlighting of the discrepancies between the negotiation routines previously described in human-human interactions and those observed in the learner-system interactions in the present study.

Type

Following these considerations, the utterance after the response was assigned the main code *reaction to response* and one of the subcodes listed in Table 20. As

indicated by the subheadings in Table 20, one pair of reaction types had the function of indicating that the non-understanding was resolved while two other pair of reaction types indicated that the non-understanding persisted. In each pair, one reaction type involved an explicit statement (i.e., *explicit resolution*, *explicit non-resolution*) while the other one involved an implicit statement (i.e., *implicit resolution*, *implicit no-resolution*). A fourth reaction type involved no reaction to response at all. *No reaction to response* typically occurred in the context where there was already *no response* to an indicator since the task had formally been completed.

This can be seen in the example in Table 20: the learner confirmed that they did not need “anything else” as the task was completed at this point of the interaction. As can be seen further, this learner message triggered an indicator of the *inappropriate response* type by the dialog system. However, the learner’s response to the indicator is absent, presumably because the learner considered it unnecessary to negotiate meaning further as the task was already complete. Consequently, there was no reaction to response by the interlocutor.

As the frequency counts in Table 20 show, *implicit resolution* was the most frequent reaction type (92 counts or 42.2%), followed by *explicit no resolution* (63 or 28.9%) and *implicit no resolution* (56 or 25.7%). *No reaction to response* was relatively infrequent (seven counts or 3.2%). As mentioned above, the *explicit resolution* reaction type, which corresponds to the *reaction to response* as previously defined in human-human interaction literature, was absent from the learner-system data but is included here for consistency and relevance.

Table 20

Frequency and Distribution of Subcodes for Type of Reaction to Response of Negotiation of Meaning in Learner-System Interaction

Subcode	Definition	Example	Count	%
Reaction to response indicating that the non-understanding has been resolved				
Implicit resolution	The interlocutor implicitly indicates that the non-understanding has been resolved by providing an appropriate	Learner: Okay and how much does it costs? [trigger] System: So sorry, I could not understand your last message. Could you please say it again? [indicator]	92	42.2

Subcode	Definition	Example	Count	%
	reaction to the preceding response move.	Learner: How much does the Alpha room costs? [response] System: <u>The cost for Alpha would be CHF 1400.</u> [reaction to response]		
Explicit resolution	The interlocutor explicitly indicates that the non-understanding has been resolved.	-	0	0.0
Reaction to response indicating that the non-understanding has not been resolved				
Implicit no resolution	The interlocutor indicates implicitly that the non-understanding has not been resolved by giving an inappropriate reaction to the preceding response.	Learner: ohh thats nice [trigger] System: Excuse me, can you please rephrase your last message? [indicator] Learner: sorry [response] System: <u>Could you please tell me your name?</u> [reaction to response]	56	25.7
Explicit no resolution	The interlocutor states explicitly that the non-understanding has not been resolved.	Learner: There are some more points [trigger] System: Please could you tell me who I am chatting with? [indicator] Learner: This is Mithusha [response] System: <u>I am sorry, I did not understand that. Could you please try one more time?</u> [reaction to response]	63	28.9
No reaction to response				
No reaction to response	There is no reaction to response because the negotiation takes place at the end of the interaction or there is a message overlap.	System: You are very welcome. System: Could I help you with anything else? Learner: No thats all. [trigger] System: The cost for Alpha would be CHF 1400. System: Room Beta comes at a cost of CHF 900. This includes the price for the terrace. System: Room Gamma would be CHF 1100. [indicator] [No response; no reaction to response / end of interaction]	7	3.2

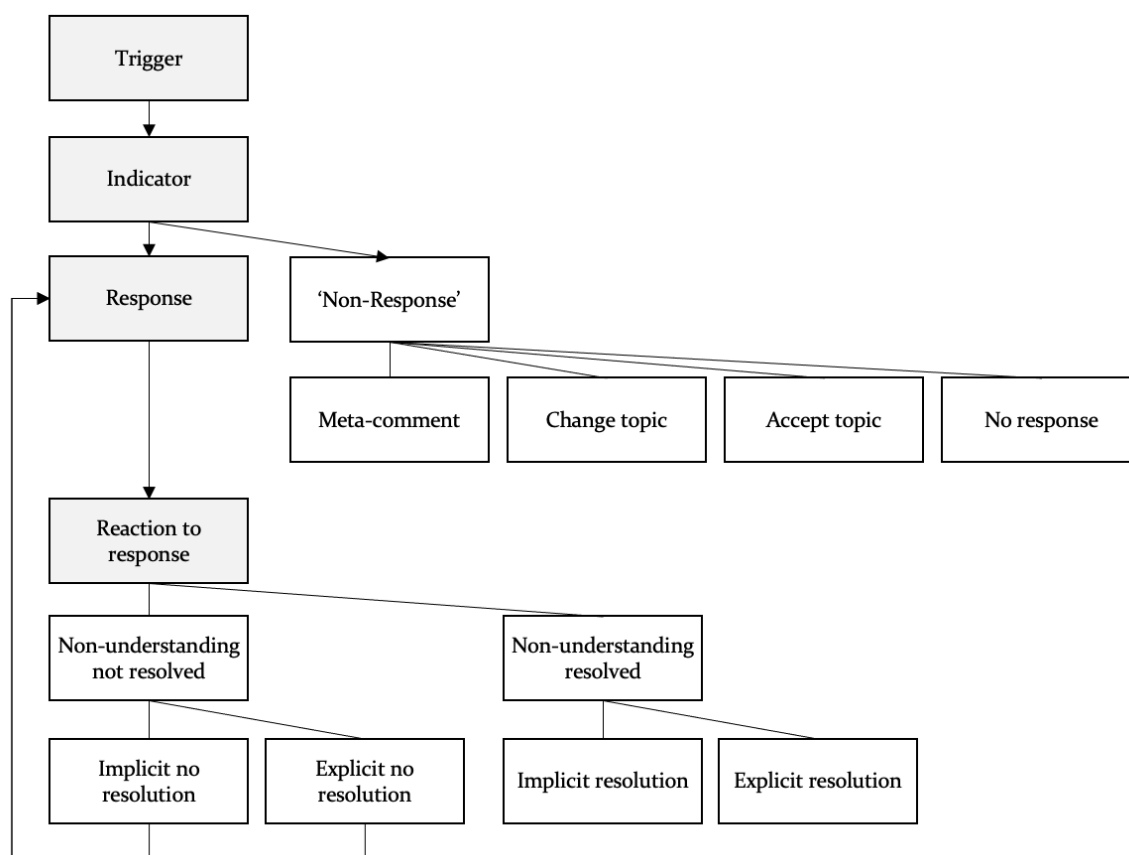
Complexity of Negotiation of Meaning Episode

As indicated in the previous paragraphs, responses such as elaboration, modification, reduction, or repetition were sometimes unsuccessful in resolving the non-understanding. As stated in the paragraph on the varied success of the different response types (i.e., elaboration, modification, reduction, repetition), responses across all four types had an overall success rate of 46.5% of resolving non-understanding. In other words, in 53.5% of the instances where non-understanding was indicated, it could not be resolved by the respondent.

Such unresolved non-understanding would often lead to another negotiation routine. Figure 15 visualizes the possible paths following responses in learner-system negotiation of meaning sequences: As shown, a response may either resolve a non-understanding immediately, which led to a reaction of response of the type *implicit resolution* (see example in Table 20) or it may not resolve the non-understanding, which led to a reaction of response of the type *implicit no resolution* or *explicit no resolution* (see examples in Table 20). In the latter case, the response simultaneously functioned as a new trigger provoking another indicator, which would in turn set off another response or, alternatively, one of the 'non-response' strategies described above and so on.

Figure 15

Visualization of Possible Paths Following the Response Component in Negotiation of Meaning in Learner-System Interaction



Such loop-backs may create negotiation routines with two or more layers of trigger-indicator-response sequences. Example 64 and Example 65 illustrate negotiation of meaning routines with two and three, respectively, layers of trigger-indicator-response (-reaction to response) sequences.

Example 64

Negotiation of meaning routine with two layers of trigger-indicator-response-reaction to response sequences

36	Learner	yeah sure how many cash would it be?	<i>Trigger 1</i>
37	System	I am pleased to hear that you like it.	<i>Indicator 1</i>
38	System	Could I be of assistance with anything else?	

39	Learner	Yes how many would it cost?	<i>Response 1 / Trigger 2</i>
40	System	I am glad you like it.	
41	System	Is there anything else I can help you with?	<i>Reaction to response 1 (Implicit no resolution) / Indicator 2</i>
42	Learner	I need to know how much would it cost	<i>Response 2</i>
43	System	The price for Gamma is 1100 CHF.	<i>Reaction to response 2 (Implicit resolution)</i>

Example 65

Negotiation of meaning routine with three layers of trigger-indicator-response-reaction to response sequences

54	Learner	I confirm	<i>Trigger 1</i>
55	System	I am sorry, I did not understand that. Could you please try one more time?	<i>Indicator 1</i>
57	Learner	I take room Alpha for the 12th october 2019.	<i>Response 1 / Trigger 2</i>
58	System	So, I will book room Alpha on the 12th october 2019 for you. Could you please confirm if you would like me to proceed?	<i>Reaction to response 1 (Implicit no resolution) / Indicator 2</i>
60	Learner	I confirm.	<i>Response 2 / Trigger 3</i>
61	System	Excuse me, could you rephrase this please?	<i>Reaction to response 2 (Explicit no resolution) / Indicator 3</i>
62	Learner	I confirm this reservation.	<i>Response 3</i>
63	System	Room Alpha has just been reserved for you.	<i>Reaction to response 3 (Implicit resolution)</i>

Multi-layered negotiation routines may either ultimately resolve the non-understanding. This was the case when the final indicator was followed by a response that resolved the non-understanding, leading to a reaction to response of the type *implicit resolution*, as illustrated in Example 64 and Example 65.

Alternatively, the final indicator may be followed by a ‘non-response’ such as meta-comment, topic change, accept topic, or no response exiting the negotiation sequence. In these instances, the non-understanding remained unresolved after two or even three responses attempting to resolve it. Example 66, Example 67, and Example 68 illustrate multi-layered negotiation routines that ended with a ‘non-response’ and were thus unsuccessful in resolving the non-understanding.

Example 66

Unresolved negotiation of meaning routine with two layers of trigger-indicator-response-reaction to response sequences

6	Learner	Yes i already looked up what rooms you've got.	<i>Trigger 1</i>
7	System	Excuse me, could you rephrase this please?	<i>Indicator 1</i>
8	Learner	I have visited your web page and saw the rooms.	<i>Response 1 / Trigger 2</i>
9	System	Room Gamma has two moveable monitors and speakers.	<i>Reaction to response 1 (Implicit no resolution)/ Indicator 2</i>
10	Learner	How much space has it? Can 150 people fit in there?	<i>'Non-response' 2 (Accept topic)</i>

Example 67

Unresolved negotiation of meaning routine with three layers of trigger-indicator-response-reaction to response sequences

17	Learner	In which room can i make also private speaks	<i>Trigger 1</i>
18	System	I am sorry, I did not understand that. Could you please try one more time?	<i>Indicator 1</i>

19	Learner	which room has the most place to have conversations	<i>Response 1 / Trigger 2</i>
20	System	Would you like me to book room None for the None?	<i>Reaction to response 1 (Implicit no resolution) / Indicator 2</i>
21	Learner	no, which room is the best for have also place for tidy conversation	<i>Response 2 / Trigger 3</i>
22	System	All three rooms are supplied with the necessary equipment to hold a presentation.	<i>Reaction to response 2 (Implicit no resolution) / Indicator 3</i>
23	Learner	in which room can i have a good atmosphere?	<i>'Non-response' 3 (Change topic)</i>

Example 68

Unresolved negotiation of meaning routine with four layers of trigger-indicator-response-reaction to response sequences

12	Learner	Please give my a answer to my question. Whats your Name?	<i>Trigger 1</i>
13	System	I am sorry, I did not understand that. Could you please try one more time?	<i>Indicator 1</i>
14	Learner	Your Name please	<i>Response 1 / Trigger 2</i>
15	System	Please could you give me your name?	<i>Reaction to response 1 (Implicit no resolution) / Indicator 2</i>
16	Learner	Yes, my Name is Harald Lesch and your name?	<i>Response 2 / Trigger 3</i>
17	System	So sorry, I could not understand your last message. Could you please say it again?	<i>Reaction to response 2 (Explicit no resolution) / Indicator 3</i>
18	Learner	Your Name is?	<i>Response 3 / Trigger 4</i>

19	System	Room Alpha has space for up to 270 people.	<i>Reaction to response 3 (Implicit no resolution) / Indicator 4</i>
20	Learner	Look.. i need a Room wih space for up 150 People	<i>'Non-response' 4 (Change topic)</i>

Table 21 provides an overview of the number of multi-layered negotiation routines leading to a resolution of the non-understanding vs. the number of routines that did not. For the purpose of comparison and completeness, the table also contains the numbers for the single-layered routines consisting of single trigger-indicator-response sequences.

As shown, there were total 139 single-layered negotiation routines, of which 33 (23.7%) led to a response and from there to a resolution of the non-understanding, whereas 106 (76.3%) were followed by a 'non-response' such as meta-comment, topic change, accept topic, or no response. Of the 23 two-layered sequences, six (26.1%) resulted in resolution whereas 17 (73.9%) did not. Of the seven three-layered sequences, one (14.3%) resulted in resolution, while six (85.7%) did not. In none of the three four-layered sequences the non-understanding was resolved. Overall, out of 172 negotiation episodes, learners and dialog system managed to resolve the non-understanding of the trigger utterance in 40 cases (23.3%), whereas in 132 (76.7%) the non-understanding remained unresolved.

As a particular note, of the four negotiation routines that were initiated by a learner and responded to by the dialog system, one was resolved while three remained unresolved.

Table 21

Frequency and Distribution of Resolved and Abandoned Sequences by Number of Layers of Negotiation of Meaning in Learner-System Interaction

Number of layers	Resolved (response after final indicator)		Abandoned (‘non-response’ after final indicator)	
	Count	%	Count	%
Single-layered	33	23.7	106	76.3
2-layered	6	26.1	17	73.9
3-layered	1	14.3	6	85.7
4-layered	0	0.0	3	100.0
Total	40	23.3	132	76.7

Negotiation of Meaning in SCMC

The following section presents the findings on negotiation of meaning in learner-peer SCMC. As described in the chapter Methods (see section General Coding Procedure), the codes to analyze the interactional features were initially created based on the learner-system interaction and then slightly adapted to match the SCMC interactions. This allowed for the identification of differences in the realization and frequency of each code between the two interaction types and, from there, highlight the unique characteristics of learner-system interactional features. Thus, the following findings on negotiation of meaning in SCMC include the same codes (if present in the data) as described in the section above on learner-system interaction, focusing on contrasting the realization and frequency of each code in the two interaction types.

Indicator

Compared to learner-system interaction, SCMC led to very few indicators of non-understanding. Table 22 provides an overview of the frequency of indicators in the two interaction types. As shown, SCMC yielded a total of five indicators. Corrected for number of interactions and text length, SCMC led to 0.28 indicators per interaction and 0.01 indicators per message. Comparing this frequency to system interaction (218 indicators overall, 10.38 indicators per interaction, 0.16 indicators per message), SCMC

produced 44 times fewer indications of non-understanding overall or 16 times fewer when corrected for number of interactions and messages.

Table 22

Frequency of Negotiation of Meaning Indicators in Learner-System Interaction and SCMC

Measure	Interaction type	
	System	SCMC
Total indicators	218	5
Indicators per interaction	10.38	0.28
Indicators per message	0.19	0.01

Note. Number of learner-system interactions $n = 21$, number of SCMC interactions $n = 18$, number of messages in learner-system interaction $n = 1139$, number of messages in SCMC $n = 611$.

It needs to be pointed out that in SCMC, an additional four utterances may appear like inappropriate responses at first glance. Upon closer analysis, however, it becomes evident that they were caused by message overlap rather than non-understanding. Therefore, they were not included in the count. Example 69 illustrates this kind of inappropriate response caused by message overlap.

Example 69

Inappropriate response cause by message overlap

11	Peer	We have also the room Beta. This have place for 130 people.	
12	Learner	I hope the room is nice decorated but it doesn't have to be too much luxury	
13	Peer	But maybe thats to small?	<i>Message overlap</i>

In comparison to system interaction, where out of 218 all but four indicators were produced by the interlocutor, i.e., the dialog system, in SCMC, the signaling of non-understanding was more equally distributed between learner and interlocutor,

i.e., peer (see Table 23). It must be pointed out, however, that the observed number of indicators was too low to be conclusive.

Table 23

Frequency and Distribution of Subcodes for Type of Indicator of Negotiation of Meaning in Learner-System Interaction and SCMC

Subcode	Definition	System (n = 21)		SCMC (n = 18)	
		Count	%	Count	%
Initiator					
Learner	The learner initiates the negotiation sequence, i.e., indicates non-understanding.	4	18.8	2	11.1
Interlocutor	The interlocutor initiates the negotiation sequence, i.e., indicates non-understanding.	214	98.2	3	16.7

Comparing the nature of the indicators between the two interaction types, in SCMC, non-understanding was signaled using the same indicator types as in learner-system interaction. Table 24 lists the indicator subtypes found in SCMC along with examples from the SCMC data and frequency counts.

Table 24

Frequency and Distribution of Subcodes for Type of Indicator of Negotiation of Meaning in Learner-System Interaction and SCMC

Subcode	System (n = 21)		SCMC (n = 18)	
	Count	%	Count	%
Type				
Clarification request	101	46.3	1	5.6
Confirmation check	8	3.7	2	11.1
Inappropriate response	109	50.0	2	11.1

While the counts in SCMC are too low to reveal clear distribution patterns, the existing instances may be analyzed qualitatively as follows: As in most indicators found in learner-system interaction, the indicator of the type *clarification request* found in SCMC indicated non-understanding on the entire content of the previous message. With respect to the confirmation checks found in SCMC, they both involved the testing of hypotheses, as in the learner-produced confirmation check found in the learner-system interaction data. It needs to be pointed out that in both, the confirmation checks as well as the clarification request, the learners in SCMC sought to clarify their roles in the task – in two instances by means of the linguistically little sophisticated German “hä” – rather than to clear up a content-related non-understanding, as was generally the case in the learner-system interactions. Two additional indicators in SCMC were of the type *inappropriate response*, which in learner-system interaction was the most frequently found indicator type.

As a note, one SCMC sequence (see Example 70) contains two overlapping negotiation episodes: The first negotiation episode is triggered by the peer’s message “its your job to book it” which created non-understanding on the side of the learner. While the peer in the role of the hotel employee understood that it was the learner’s, i.e., client’s, job to book the room, the learner, i.e., client, insisted that it was the peer’s, i.e., hotel employee’s, task. The second negotiation episode involved the same disagreement on task roles: The learner expresses puzzlement (“hä”) in response to the interlocutor’s statement “its your job to book”. The learner’s confused “hä” triggers non-understanding on the side of the interlocutor who produces the indicator “not?” to confirm whether they had in fact misunderstood the roles.

Example 70

Two overlapping episodes of negotiation of meaning in SCMC

58	Peer	its your job to book it	<i>Trigger 1</i>
59	Learner	Hä	<i>Indicator 1 (Clarification request) / Trigger 2</i>
60	Peer	not?	<i>Response 1 / Indicator 2 (Confirmation check)</i>
61	Learner	No	

62	Learner	you must book the roomfor me	<i>Reaction to response 1 / Response 2</i>
63	Peer	its my job? Okey	
64	Peer	pipipi its booked	<i>Reaction to response 2</i>

Trigger

As in learner-system interaction, the triggers found in SCMC could not be classified further according to the specific element that caused non-understanding. As the examples above showed, all indicators related the non-understanding to the entire content of the previous message or, in the case of the inappropriate responses, contained no indication at all. It also should be noted that three out of the five non-understandings were related to the task rather than content, more precisely, the responsibilities of the two roles client and hotel employee.

Furthermore, as in learner-system interaction, in SCMC both erroneous and target-like utterances led to indications of non-understanding. Table 25 shows an overview of the number and percentage of correct and erroneous triggers leading to indications of non-understanding in the two interaction types. As can be seen, SCMC led to a comparable number of 'false-negatives', i.e., indications of non-understanding in response to triggers that are target-like, as learner-system interaction.

Table 25

Frequency and Distribution of Subcodes for Accuracy of Trigger of Negotiation of Meaning in Learner-System Interaction and SCMC

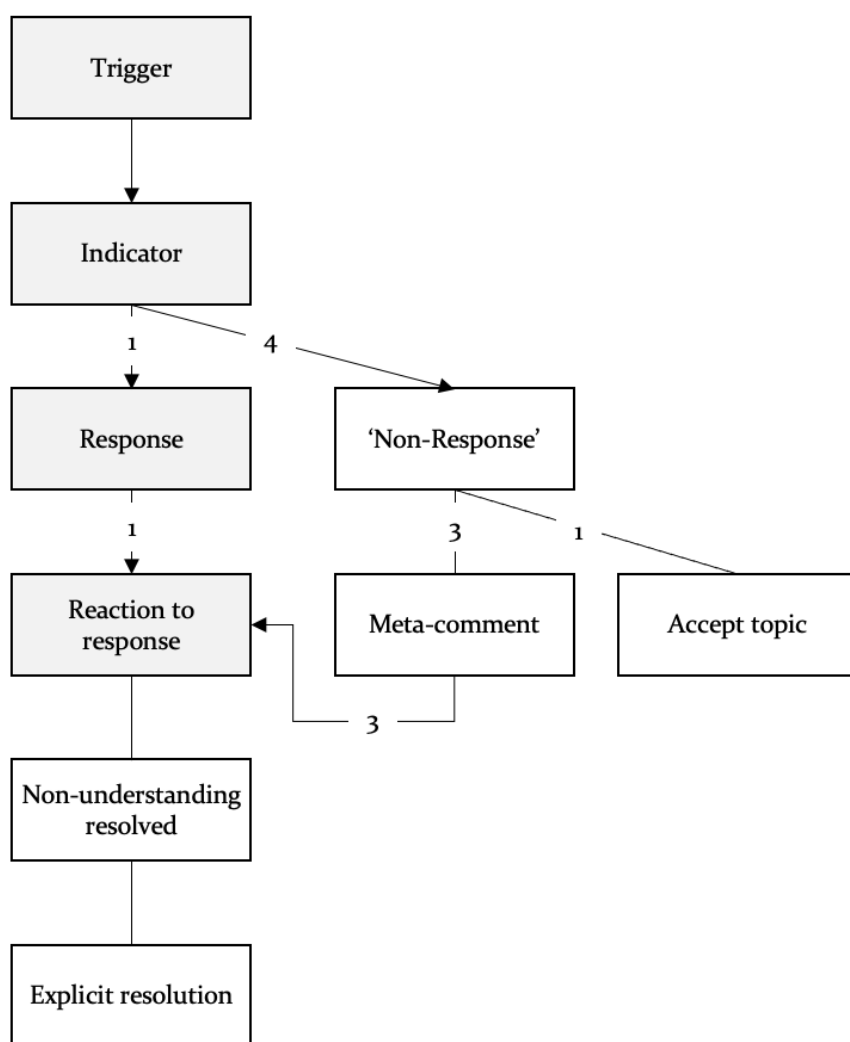
Subcode	System (<i>n</i> = 21)		SCMC (<i>n</i> = 18)	
	Count	%	Count	%
Accuracy				
Erroneous	121	55.5	3	60.0
Target-like	97	44.5	2	40.0

Response

Another parallel between the two interaction types concerns the strategies used in responding to the indicator. The visualization in Figure 16 describes the paths the learners and peer interlocutors may follow in SCMC. It also indicates the frequencies with which each route was followed.

Figure 16

Visualization of Possible Paths Following the Response Component in Negotiation of Meaning in SCMC



As the visualization shows, learners and interlocutors in SCMC, as in learner-system interaction, either rephrased the trigger utterance in a response or used a ‘non-response’ such as *meta-comment* or *accept topic*. With regard to the ‘non-response’

types, no instances of *no response* were found in the SCMC data, as opposed to seven instances that were found in learner-system interaction. Instead, ‘non-responses’ in SCMC exclusively involved *meta-comment* and *accept topic*. The extracts containing these two types of non-response strategies found in the SCMC data are presented in Example 71 and Example 72.

Example 71

‘Non-response’ type Meta-comment in SCMC

9	Learner	Yes, but in which hotelroom do we wanna go?	
10	Learner	it has Alpha, Beta and Gamma hotelroom	<i>Trigger</i>
11	Peer	aren't you the customer?	
12	Peer	I work in the one king hotel	<i>Indicator</i>
13	Learner	aha omg now i understand the exersice	<i>‘Non-response’ (Meta-comment)</i>

Example 72

‘Non-response’ type Accept topic in SCMC

43	Learner	is it the best room for my "vorgaben"?	<i>Trigger</i>
44	Peer	Great	<i>Indicator</i>
45	Learner	Okey	
46	Learner	i want to book it	<i>‘Non-response’ (Accept topic)</i>

Importantly, as the visualization in Figure 16 also indicates, meta-comments in SCMC did not interrupt the negotiation as they did in learner-system interaction. Instead, meta-comments in SCMC were productive in resolving the non-understanding, as illustrated in Example 73.

Example 73

‘Non-response’ type Meta-comment leading no resolution of non-understanding in SCMC

3	Peer	Who are you?	<i>Trigger</i>
---	------	--------------	----------------

4	Learner	Good and you?	<i>Indicator</i>
5	Peer	Not how, who?	<i>'Non-response' (Meta-comment topic)</i>
6	Learner	aha ups I'm Anna	<i>Reaction to response (Explicit resolution)</i>

As the visualization Figure 16 also shows, only a single response that acknowledged and responded to the non-understanding of the indicator occurred. The excerpt is presented in Example 74. So, while the number of responses found in SCMC was too low to conclusively compare them with the system interaction, the following observations can be made about the existing example in SCMC: the response was of the *elaboration* subtype, which was also found and realized in much the same way as in learner-system interaction, apart from the message split between “no” and “you must book [...]” which was not observed in learner-system interaction. The remaining subtypes found in learner-system interaction, i.e., *reduction*, *modification*, *repetition*, did not occur in the SCMC data.

Example 74

Response in SCMC

58	Peer	its your job to book it	
59	Learner	Hä	<i>Trigger</i>
60	Peer	not?	<i>Indicator</i>
61	Learner	No	
62	Learner	you must book the roomfor me	<i>Response (Elaboration)</i>
63	Peer	its my job? Okey	
64	Peer	pipipi its booked	<i>Reaction to response (Explicit resolution)</i>

Reaction to Response

All four SCMC negotiation sequences that were completed included a reaction to the response as per the definition generally used in human-human interaction literature quoted above, i.e., explicitly indicating that the non-understanding had been resolved, as shown in Example 73 above as well as an Example 75. This contrasted with learner-system interaction where this type never occurred and, instead, reactions to response involved either an implicit indication of resolved non-understanding or an implicit or explicit indication of continued non-understanding, depending on the success of the response.

Example 75

Reaction to response in SCMC

9	Learner	Yes, but in which hotelroom do we wanna go?	
10	Learner	it has Alpha, Beta and Gamma hotelroom	<i>Trigger</i>
11	Peer	aren't you the customer?	
12	Peer	I work in the one king hotel	<i>Indicator</i>
13	Learner	aha omg now i understand the exersice	<i>'Non-response' (Meta-comment)</i>
14	Peer	hahhah	
15	Peer	okay lets begin	
16	Peer	Which room do you want to mieten?f	<i>Reaction to response (Explicit resolution)</i>

Since in all negotiation routines found in SCMC, the response or meta-comment, respectively, led to an immediate resolution of the non-understanding, no loop-backs occurred, i.e., no multi-layered sequences, as opposed to the learner-system interactions.

Learners' Perception on Negotiation of Meaning

With regard to negotiation of meaning, the post-questionnaire included prompts on the degree of mutual understanding during the interactions and how learners dealt with non-understanding. When asked how well they understood their conversational partner, learners from both groups (system, SCMC) reported that they understood *always or almost always* (system: 15 respondents or 71.4%; SCMC: 14 respondents or 77.8%) or *often* (system: five respondents or 23.8%, SCMC: four respondents or 22.2%). In the system group, one respondent (4.8%) reported that they understood their interlocutor only *sometimes*.

When asked about the strategies the learners used when they could not understand their conversational partner, both groups tended to report using non-interactive strategies, such as *looked up online, derived meaning from context, checked in chat, or changed topic*, while the reported use of interactive strategies such as *asked interlocutor if understood correctly or told interlocutor that not understood* was low. The reported number of strategies was very low overall (system: eight reported strategies; SCMC: seven reported strategies). For an overview of the reported strategies including counts, see Appendix J.

When asked how well the conversational partner understood the learner, there was a noticeable difference between the two groups: While learners from the SCMC group reported that their conversational partner understood them *always* (15 respondents or 83.3%) or *often* (three respondents or 16.7%), more than half of the learners who interacted with the dialog system reported that their interlocutor understood them only *sometimes* (10 respondents or 47.6%) or even *rarely* (two respondents or 9.5%). The remaining learners reported that they had been understood *often* (eight respondents, 38.1%). Only one respondent (4.8%) responded that they were *always* understood.

When asked about how they reacted to non-understanding on the side of their conversational partner, learners from the system group reported a much higher use of strategies to resolve the communication issue than learners from the SCMC group (system: 49 reported strategies; SCMC: five reported strategies). The learners interacting with the dialog system also reported using a wide variety of different

strategies (for a complete overview, see Appendix J) which corresponded to the response types found in the interactional data (see section Negotiation of Meaning in the chapter Findings) such as *repeated message* (corresponding to *repetition*), *changed spelling*, *used a different word*, *changed syntax* (corresponding to *modification*), *wrote more detailed*, *summarized multiple messages into one* (corresponding to *elaboration*), *wrote less detailed*, *split statement into multiple messages* (corresponding to *reduction*), *changed topic* (corresponding to the ‘non-response’ *topic abandonment*), or *asked interlocutor if understood* (corresponding to ‘non-response’ *meta-comment*). As other strategies, one learner additionally named *confusing the interlocutor*, which presumably refers to copy-pasting the dialog system’s own messages as input, as could be observed in the respondent’s interaction. The remaining three responses given under other strategies could in fact be subsumed under the already existing strategies.

In this context, it is important to point out that three learners from the system group specifically expressed their irritation or frustration about the system’s lack of understanding, with one learner describing it as “frustrating” and “tedious”.

The interviews revealed additional findings on the learners’ perception on negotiation of meaning. An overview of the codes derived from the analysis of the interviews relating to negotiation of meaning can be found in the appendix (Appendix K). It needs to be noted that the data was exclusively gathered from interviews conducted with participants of the learner-system group since none of the SCMC participants commented on negotiation of meaning. So, the following findings reflect solely the perspective of learners who interacted with the dialog system.

The learners’ statements related to three different aspects of negotiation of meaning: the learners’ emotional reactions to the interlocutor’s non-understanding, their conceptions about the reasons for the non-understanding, and their strategies to deal with the non-understanding.

Focusing on the learners’ emotional reactions, they expressed five types of emotional reactions to the interlocutor’s non-understanding (for an overview with examples, see Appendix K). Three of the types were negative emotions, i.e., *frustrated/annoyed*, *lost*, and *confused*. The remaining two types, *curious* and *amused*, may be considered positive emotions, although amusement may also indicate that

learners did not take the interaction and/or their conversational partner seriously, which would rather be considered negative from a motivational point of view.

The code matrix in Figure 17 shows the subcodes for the aspect *emotional reaction* that were assigned to each participant's interview. As the overview shows, most of the interviewed learners experienced a range of different negative and positive emotions in reaction to the dialog system's lack of understanding, with the exception of one learner (5CT-2N9) who mentioned "confused" as the only emotional reaction.

Figure 17

Code Matrix of Subcodes for Emotional Reactions to Negotiation of Meaning per Learner in Learner-System Interaction

Code System	CW7-ZDE	KM8-5TV	357-KW8	4XE-72U	5CT-2N9	RC5-SP9
INTERACTIONAL FEATURES						
NEGOTIATION OF MEANING	■	■	■	■	■	■
Emotional reaction	■	■	■	■	■	■
Lost patience		■	■	■		■
Frustrated/Annoyed		■		■		■
Amused	■		■			
Curious	■			■		
Confused	■		■	■	■	■

Note. Only code assignments for interviews with participants from the learner-system group are displayed since comments, and hence, subcodes on negotiation of meaning were absent in interviews with participants from the SCMC group.

The coding of the interview responses further revealed that learners interacting with the dialog system had a range of conceptions explaining the interlocutor's understanding or non-understanding, respectively, of a given message (for an overview with examples, see Appendix K). The learners' conceptions can be divided into two groups: those who ascribe the reason for the non-understanding to a shortcoming of the interlocutor (e.g., *the dialog system cannot process longer messages or messages containing multiple requests, the dialog system is not capable of retaining information, i.e., does not have a 'memory', the dialog system cannot understand all words, i.e., has a limited vocabulary*) and those where the learner ascribes the reason for the non-understanding to their own production which they view as inadequate in some way (e.g., *non-understanding due to inadequate discourse management, non-understanding due to a lexical error such as lexical misselection or use of German*).

As explained above (see section Dialog System), the dialog system's non-understanding may be caused by spelling error detection or incorrect intent assignment. In turn, incorrect intent classifications occurred when the dialog system was confronted with wordings that were not included in the training data or because different intents were similarly worded, and the dialog manager was at the same time unable to mitigate the effect considering the context of the learner message. Another source of error are intents that were unknown to the model, either because they were out of scope (e.g., "Which one of the rooms has been the most popular up until now?") or when a message included several intents.

So, the learners correctly conceived that both their own interactional behavior as well as the dialog system played a role in the dialog system's non-understanding. Learners were also correct in assuming that the dialog system was often unable to process messages containing multiple intents, or linguistic input ("vocabulary") that was not included in the training data.

The remaining conceptions were partly true: Concerning the retaining of information, the dialog system did have the capability of storing some information. This capability was, however, limited to predefined entities such as the user's name, budget, booking date etc. In addition, preformulated 'generic' responses may give the impression that the dialog system did not customize the response to specifically formulated requests and thus disregarded what the user may have said before. Similarly, the conception that the dialog system only extracted keywords from a user's message is a simplified view. While keyword extraction did play a role in the NLU's mechanism, the dialog system also relied on sentence structure as well as the context of a message within the dialog. The latter is also the reason why a learner was under the impression that the dialog system was sometimes unable to understand the first time. Indeed, an identical message might not be accurately classified in a first instance (*first time*) but when it is repeated due to the different context.

Turning to the learners' reasonings concerning their own role in the non-understandings, learners' often could not identify what aspect of their production might have caused a problem (coded as *unspecified*). When specifying the aspect that might have led to non-understanding, they most frequently mentioned *lexical errors* and, more rarely, *inadequate discourse management* or *unclear reference*.

The visualization in Figure 18 displays the subcodes for conceptions that were assigned to each learner interview. As shown, most learners established at least two different conceptions and saw themselves as well as the dialog system responsible for the non-understanding. An exception was one participant (KM8-5TV) who ascribed no fault to the dialog system.

Figure 18

Code Matrix of Subcodes for Conceptions on Negotiation of Meaning per Learner in Learner-System Interaction

Code System	CW7-ZDE	KM8-5TV	357-KW8	4XE-72U	5CT-2N9	RC5-SP9
INTERACTIONAL FEATURES						
NEGOTIATION OF MEANING						
Conceptions						
Interlocutor's fault						
Long message	•					
First time			•			•
Retaining of information			•	•		
Vocabulary				•	•	
Keyword-based						•
Learner's fault						
Unspecified	•	•	•	•	•	
Inadequate discourse m	•		•			
Lexical error		•	•			•
Unclear reference			•			•

Note. Only code assignments for interviews with participants from the learner-system group are displayed since comments, and hence, subcodes on negotiation of meaning were absent in interviews with participants from the SCMC group.

Furthermore, the interviewed learners explained what strategies they used in reaction to the interlocutor's indication of non-understanding (for an overview of the reported strategies with examples, see Appendix K). The strategies included discourse management strategies such as restarting the entire conversation, repeating a message, or switching from packing multiple requests into a single message to stating one request per message – or the other way around. Other strategies such as *keyword* and *rephrase* occurred at the level of the message.

The learners' strategies also included the conscious decision of giving up, i.e., not attempting to resolve the non-understanding and instead changing the topic or leaving the interaction. Closer analysis of the text segments coded as *give up* indicated

that learners chose this strategy because they were frustrated or annoyed with the dialog system's non-understanding or considered the system utterance irrelevant.

The visualization in Figure 19 shows which conceptions and strategies each interview participant mentioned. As can be seen in this overview, the learners' conceptions were generally met by a related strategy, although it needs to be pointed out that the conceptions here only reflect those mentioned by the learners during the interview and learners might have used further strategies without mentioning them.

One learner (CW7-ZDE), for instance, conceived that non-understanding was a result of their own inadequate discourse management and the interlocutor's lack of ability to process long messages. Accordingly, the learner stated that they proceeded step-by-step, i.e., sending messages with single pieces of information.

Similarly, two other learners (KM8-5TV and 5CT-2N9) suspected that their lexical choices led to non-understanding and focused on rephrasing and the use of a single keyword. The further strategies, *meta-comment* and *restart*, respectively, both reflect that the learners had no clear idea what element of their production had caused non-understanding.

The overview also shows that all but one learner had at some point unspecified conceptions about the shortcoming of their own production and mentioned *give up* as a strategy that they had used during episodes of non-understanding.

Figure 19

Code Matrix of Subcodes for Strategies and Conceptions on Negotiation of Meaning per Learner

Code System	CW7-ZDE	KM8-5TV	357-KW8	4XE-72U	5CT-2N9	RC5-SP9
INTERACTIVE FEATURES						
NEGOTIATION OF MEANING						
Conceptions						
Interlocutor's fault						
Long message	●					
First time			●			●
Retaining of information			●	●		
Vocabulary				●	●	
Keyword-based						●
Learner's fault						
Unspecified	●	●	●	●	●	
Inadequate discourse m	●		●			
Lexical error		●	●			●
Unclear reference			●			●
Strategies						
Restart					●	
Repeat						●
Meta-comment		●				●
Keyword					●	
All-in-one				●		
Step-by-step	●		●			
Rephrase		●	●		●	
Give up		●	●	●	●	●

Note. Only code assignments for interviews with participants from the learner-system group are displayed since comments, and hence, subcodes on negotiation of meaning were absent in interviews with participants from the SCMC group.

4.2.2 Recasts

As presented above (see section Feedback in the chapter Literature Review), recasts have been described in the literature as a type of implicit feedback by presenting the learner with a more target-like reformulation of their incorrect utterance while maintaining its original meaning (Loewen, 2011). As stated in the chapter Methods (see section Coding of Recasts), the present study defined recasts as episodes in which the interlocutor rephrases the learner's preceding utterance that contains at least one error to be more target-like by changing one or more sentence components while retaining its central meaning and without breaking the flow of the communication.

Recasts in Learner-System Interaction

In learner-system interaction a total of 40 recasts or 1.90 recasts per interaction and 0.04 per message were found, as shown in Table 26.

Table 26

Frequency of Recasts in Learner-System Interaction

Measure	Frequency
Total recasts	40
Recasts per interaction	1.90
Indicators per message	0.04

Note. Number of learner-system interactions $n = 21$, number of messages in learner-system interactions $n = 1139$.

As detailed in the chapter Methods (see section Coding of Recasts), the two components of the recast sequence, i.e., recast move and uptake, were further classified into subcodes relating to their linguistic focus and efficiency in functioning as feedback. An overview of all recast subcodes is presented again in Table 27. The subcodes are introduced in detail with examples and frequency counts in the following paragraphs.

Table 27

Overview of Main and Subcodes for Recasts With Counts

Main code	Subcode	Count	
		System ($n = 21$)	SCMC ($n = 18$)
Recast move			
Linguistic focus	Spelling	26	3
	Lexis	6	-
	Grammar	6	6
	Multi-focus	2	2
Intensity of recast	Singular incidence	23	11
	Repeated incidence	17	-
Intensity of recast form	Singular incidence	34	8
	Repeated incidence	6	3

Main code	Subcode	Count	
		System (n = 21)	SCMC (n = 18)
Completeness	Complete	18	-
	Partial	22	11
Length	Single sentence	19	1
	Multiple sentences	11	2
	Multiple messages	10	8
Uptake	No uptake	20	9
	Immediate uptake		
	Repair	3	-
	Needs repair	1	1
	Delayed uptake		
	Repair	5	1
	Needs repair	11	-
	'Re-uptake'		
	Repair	3	-
Needs repair	8	-	

Linguistic Focus

A first group of subcodes relate to the linguistic focus of the recast, i.e., the linguistic aspect of the error that the recast addressed. Recasts found in the system interactions focused on grammar, spelling, or lexis, or on multiple error types simultaneously.

Table 28 presents an overview of the four subcodes pertaining to the linguistic focus of recasts along with definitions, examples, and counts for each subcode. Further examples and observations on each subcode will be presented below. As shown in Table 28, the majority of recasts (26 recasts or 65.0%) addressed grammatical learner errors. Six recasts (15.0%) focused on spelling, and another six (15.0%) on lexis. Two recasts (5.0%) addressed multiple error types, with both recasts simultaneously focusing on the error types *grammar* and *spelling*. Note that for better clarity in all following examples only those learner errors that are involved in the recast are marked in the learner utterance.

Table 28

Frequency and Distribution of Subcodes for Type of Linguistic Focus of Recasts in Learner-System Interaction

Subcode	Definition	Example	Count	%
Grammar	The recast addresses one or more grammatical errors in the learner's utterance.	Learner: Look.. i need a Room with space for up <u>150</u> [grammar error] People System: There are three rooms available for 150 people. Room Alpha has space for <u>up to 270</u> [recast form] people. Room Beta has space for 130 guests and a terrace for an additional 60 people.	26 ^a	65.0
Spelling	The recast addresses one or more spelling errors in the learner's utterance.	Learner: I'd like to take one of this 3 rooms, Alpha, Beta or <u>Gamm</u> [spelling error], wich one is the Best room and how much does it cost ? System: The price for <u>Gamma</u> [recast form] is 110 CHF.	6	15.0
Lexis	The recast addresses one or more lexical errors in the learner's utterance.	Learner: thats right can you reservate [lexis error] the room please System: For what date would you like to <u>reserve</u> [recast form] room Gamma?	6	15.0
Multi-focus	The recast addresses more than one error type in the learner's utterance.	Learner: I was asking if the room <u>Betta</u> [error 1 (spelling)] <u>have</u> [error 2 (grammar)] a <u>Projecter</u> [error 3 (spelling)], we won't to show are Gest a Video ? System: Room Alpha offers a stage, state-of-the-art equipment including a large screen, a <u>projector</u> [recast form on error 3] and dolby sound system. Room <u>Beta</u> [target-like form error 1] <u>has</u> [recast form on error 2] a portable screen, a <u>projector</u> [recast form on error 3], and loudspeakers. Bot: Gamma has two big portable screens and loudspeakers.	2	5.0

^a Of the 26 *grammar* recasts, 21 occurrences involved the superfluous use of the definite article *the* with room names (e.g., "in the room Beta")

Grammar

As a note on grammar recasts, 21 of the 26 found instances concerned the superfluous use of the definite article *the* with room names, as illustrated in Example 76.

Example 76

Grammar recast on the superfluous use of the definite article the with room names in learner-system interaction

44	Learner	i take <u>the room gamma</u>	<i>Error</i>
45	System	Yes, absolutely.	
46	System	For what date would you like me to reserve <u>room Gamma</u> for you?	<i>Recast form</i>

The remaining grammar recasts addressed a variety of grammatical errors such as the omission of prepositions or indefinite articles (Example 77), the use of the indefinite article *an* instead of *a* for a word starting with a consonant (Example 78), or the incorrect inflection for person of nouns (Example 79) and verbs (Example 80).

Example 77

Grammar recast on the omission of the indefinite article an in learner-system interaction

18	Learner	Yes sure, I would like to book a room for minimum 150 people in your hotel. The room should let us do a presentation and also meeting for small groups. It should also have a good atmosphere. It should make the guests happy. Do you have <u>room</u> for maximum 1500 Franks?	<i>Error</i>
19	System	Yes, certainly.	
20	System	Thank you for contacting One King Hotel. I will gladly assist you in finding <u>a room</u> .	<i>Recast form</i>

Example 78

Grammar recast on the incorrect use of the indefinite article an instead of a in learner-system interaction

- | | | | |
|----|---------|---|--------------------|
| 26 | Learner | Have you got a chance to do <u>an</u> presentation in the Alpha room? | <i>Error</i> |
| 27 | System | Room Alpha is equipped with a big screen, <u>a</u> sound and lighting system as well as <u>a</u> stage | <i>Recast form</i> |

Example 79

Grammar recast on the incorrect inflection for person of nouns in learner-system interaction

- | | | | |
|----|---------|--|--------------------|
| 10 | Learner | Have you got any rooms with space for at least 150 <u>Peoples</u> ? | <i>Error</i> |
| 11 | System | Yes, certainly. | |
| 12 | System | We have three rooms that are suitable for None guests. | |
| 13 | System | Room Alpha has space for up to 270 <u>people</u> . | <i>Recast form</i> |
| 14 | System | The Beta room has space for a total of 190 <u>people</u> , indoor and outdoor. | <i>Recast form</i> |
| 15 | System | There is space for 200 <u>people</u> in room Gamma. | <i>Recast form</i> |

Example 80

Grammar recast on the incorrect inflection for person of verbs in learner-system interaction

- | | | | |
|----|---------|--|--------------------|
| 17 | Learner | Which room <u>have</u> the best atmosphere? | <i>Error</i> |
| 18 | System | Room Alpha is a spacious auditorium on the basement floor with comfortable seating and a stage. It is equipped with multimedia facilities and a lighting system. | |
| 19 | System | Beta <u>has</u> a modern touch. It is located on the 15th floor and offers incredible views. The | <i>Recast form</i> |

		glass roof brings in natural light and creates a beautiful atmosphere.	
20	System	Room Gamma <u>has</u> the original decoration from the 1920s which gives the room an elegant look.	<i>Recast form</i>

Spelling

Another group of recasts focused on spelling errors, as illustrated in the following examples. Recasts focusing on spelling involved a variety of error types such as for example the misselection or omission of letters (Example 81) as well as ‘mechanical’ typing errors (Example 82)

Example 81

Spelling recast on the omission of a letter in learner-system interaction

23	Learner	But i will book a room, <u>wich</u> room do you think is the best for my options?	<i>Error</i>
24	System	Yes, sure.	
25	System	Could you please tell me <u>which</u> room you are referring to?	<i>Recast form</i>

Example 82

Spelling recast on a typing error in learner-system interaction

20	Learner	I hereby confirm, that I would like to book room Alpha for <u>thw</u> 12th of October 2019.	<i>Error</i>
21	System	So, I will book room Alpha on <u>the</u> 12th of October 2019 for you. Could you please confirm if you would like me to proceed?	<i>Recast form</i>

Lexis

The examples below illustrate recasts focusing on lexical errors. Recasts focusing on lexis addressed errors involving the semantically inappropriate choice of lexical items (as seen in Example 83), the invention of non-existing lexical items (Example 84), or the use of German (Example 85).

Example 83

Lexical recast on the semantically inappropriate choice of a lexical item in learner-system interaction

17	Learner	We want the room on the 12. octobre 2019 <u>reservation</u> , but how many costs this room?	<i>Error</i>
18	System	Would you like me to <u>book</u> room Alpha for the 12 octobre 2019?	<i>Recast form</i>

Example 84

Lexical recast on the invention of a non-existing lexical item in learner-system interaction

28	Learner	Yes how much is the <u>costen</u> for this room?	<i>Error</i>
29	System	The <u>price</u> is CHF 1100 CHF for room Gamma.	<i>Recast form</i>

Example 85

Lexical recast on the use of German in learner-system interaction

13	Learner	Both. Do they have a <u>beamer</u> in?	<i>Error</i>
14	System	Alpha is equipped with a state-of-the-art multimedia set including a big screen, <u>projector</u> and sound system.	<i>Recast form</i>
15	System	Room Beta can be equipped with a moveable screen, <u>projector</u> , and loudspeakers.	<i>Recast form</i>

Intensity of Recast Within Interaction

This dimension relates to the frequency with which the same error was recast within a given interaction. As displayed in Table 29, 17 recasts (42.5%) were of the type *repeated incidence*, i.e., at least one of the linguistic forms addressed in the recast sequence was addressed again in another recast sequence within the same interaction. This is illustrated in the example in Table 29, where the same error (superfluous definite article *the* with room name) is first addressed in the recast sequence in messages 33 to 34 and again in messages 41 to 42. As a note, all errors that were repeatedly recast were superfluous uses of *the* with room name.

In contrast, 23 (57.5%) of recasts were of the type *singular incidence*, i.e., none of the erroneous items addressed by the recast were addressed again in another recast within the same interaction.

Table 29

Frequency and Distribution of Subcodes for Intensity of Recasts Within Interaction in Learner-System Interaction

Subcode	Definition	Example	Count	%
Singular incidence	None of the erroneous linguistic items that are addressed in the recast sequence are addressed again in any other recast sequence within the interaction.	[Message 20] Learner: I hereby confirm, that I would like to book room Alpha for <u>thw</u> [error] 12th of October 2019. [Message 21] System: So, I will book room Alpha on <u>the</u> [recast form] 12th of October 2019 for you. Could you please confirm if you would like me to proceed? [Recast (Singular incidence)] [Messages 22 – 26] [...] [End of interaction]	23	57.5
Repeated incidence	At least one erroneous linguistic item that is addressed in the recast sequence is also addressed in another recast sequence within the interaction.	[Message 33] Learner: and how is <u>the</u> [error] room alpha? [Message 34] System: <u>Room Alpha</u> [recast form] would be CHF 1400. [Recast 1] [Messages 35 – 40] [...] [Message 41] Learner: can you give me some informations about <u>the</u> [grammar error] room Beta [Message 42] System: <u>Room Beta</u> [recast form] is on the 15th floor and oversees the entire city. It has a modern ambience and connects to a lovely outside terrace. [Recast 2 (Repeated incidence)]	17	42.5

Intensity of Recast Form Within Recast

Similarly, the subcodes subsumed under *intensity of recast form within recast* describe the frequency with which the target-like form, i.e., the recast form, occurs within a given recast move. *Repeated incidence* refers to instances where the recast

form occurs more than once in the same recast move. Recasts coded as *singular incidence* provide the recast form only once within the same recast move. As Table 30 indicates, the majority of recasts (34 counts or 85.0%) contained a singular incidence of recast form, while six recasts (15.0%) contained at least one repetition of the recast form within the same recast move.

Table 30

Frequency and Distribution of Subcodes for Incidence of Target-Like Form Within Recast in Learner-System Interaction

Subcode	Definition	Example	Count	%
Singular incidence	None of the target-like forms occur more than once in the same recast move.	Learner: and what about <u>the</u> [error] Alpha? System: <u>Room Alpha</u> [recast form] is a spacious room located on the basement floor of the hotel. It is an auditorium furnished with a stage, rows of seats, and a multimedia and lighting system.	34	85.0
Repeated incidence	At least one of the target-like forms is repeated at least once in the same recast move.	Learner: Which room <u>have</u> [error] the best atmosphere? System: Room Alpha is a spacious auditorium on the basement floor with comfortable seating and a stage. It is equipped with multimedia facilities and a lighting system. Beta <u>has</u> [recast form 1] a modern touch. It is located on the 15 th floor and offers incredible views. The glass roof brings in natural light and creates a beautiful atmosphere. Room Gamma <u>has</u> [recast form 2 (Repeated incidence)] the original decoration from the 1920s which gives the room an elegant look.	6	15.0

Completeness

The subcodes relating to *completeness* describe whether the recast move addressed all learner errors. The subcode *complete recast* includes recast moves where all errors that the learner committed in the previous utterances are addressed. As displayed in Table 31, 18 out of 40 recasts (45.0%) were complete. It needs to be pointed out that the example stated in Table 31 represents the only instance where a complete recast involved multiple errors. In the remaining cases, the triggering learner utterance contained a single error that was addressed in the recast move and was consequently coded as *complete recast*, as shown in Example 86.

Example 86

Complete recast in learner-system interaction

20	Learner	I would like to book <u>the</u> room Alpha	<i>Error</i>
21	System	For what date would you like me to reserve <u>room Alpha</u> for you?	<i>Recast form</i>

Partial recast describes recast moves where at least one of the learner errors of the triggering utterance was left unaddressed. Of the total 40 recasts, 22 (55.0%) were partial recasts.

Table 31

Frequency and Distribution of Subcodes for Completeness of Recasts in Learner-System Interaction

Subcode	Definition	Example	Count	%
Complete recast	Recast sequences where all errors of the learner's utterance are addressed.	Learner: how much does <u>the</u> [error 1] room beta <u>coast</u> [error 2]? System: The <u>cost</u> [target-like form on error 2] for Beta [recast form on error 1] would be CHF 900. The terrace is included in the price.	18	45.0
Partial recast	Recast sequences where one or more errors of the learner's	Learner: I am very <u>intresed</u> [error 1] in	22	55.0

Subcode	Definition	Example	Count	%
	utterance are left unaddressed.	<u>the</u> [error 2] room Alpha System: For what date would you like to reserve <u>room Alpha</u> [recast form on error 2]?		

Length

A final classification was made for the length of the recast move. The coding distinguished between the lengths *single sentence*, *multiple sentences*, and *multiple messages*. Table 32 presents the detailed definitions of each subcode along with examples and frequency counts. As seen in the overview, about half of the recast moves were realized as single sentences (19 counts, 47.5%), whereas the other half was split about equally between recast moves of multiple sentences (11 counts, 27.5%) and recast moves consisting of multiple messages (10 counts, 25.0%).

Table 32

Frequency and Distribution of Subcodes for Length of Recasts in Learner-System

Interaction

Length	Definition	Example	Count	%
Single sentence	The recast move consists of a single sentence consisting of one or more clauses.	Learner: How many <u>person</u> [error] fit in there? System: <u>There is space for 200 people</u> [recast form] <u>in room Gamma.</u>	19	47.5
Multiple sentences	The recast move consists of more than one sentence in a single message.	Learner: I think it would be <u>the</u> [error] room beta with additional 20 people System: <u>Very well. Would you like to book room Beta</u> [recast form]?	11	27.5
Multiple messages	The recast move consists of multiple messages.	Learner: Both. Do they have a <u>beamer</u> [error] in?	10	25.0

Length	Definition	Example	Count	%
		System: <u>Alpha is equipped with a state-of-the-art multimedia set including a big screen, projector [recast form] and sound system.</u>		
		Room Beta can be <u>equipped with a moveable screen, projector [recast form], and loudspeakers.</u>		
		Room Gamma has two <u>moveable monitors and speakers.</u>		

Uptake

The following section reports on the uptake, i.e., the learner's response to the recast. As a note, the operationalization of *uptake* in the present study included both responses immediately following a recast move and delayed responses that occurred later in the interaction. This is in line with Smith's (2005) work on uptake in written SCMC. It also seemed conducive as the effect of noticing and uptake in written chat can be expected to extend beyond the utterance immediately following the recast move.

Table 33 presents an overview of the different types of uptake found in the learner-system interaction data. As the overview in Table 33 shows, in exactly half of the recast sequences (20 counts, 50.0%), *no uptake* occurred, i.e., the feedback was not acknowledged by the learner and the recast form did not occur anywhere in the learner's production following the recast move.

In the other 20 cases where uptake did occur, four instances (20.0%) were coded as *immediate*, i.e., the recast form was produced by the learner in the turn immediately following the recast move. Of these, three included *repair*, i.e., the production of the target-like form, whereas one immediate uptake was of the type *needs repair*, i.e., contained the recast item with the original or a new error. The remaining 16 uptakes (80.0%) occurred in a later turn as *delayed uptake*, either *with repair* (five counts) or as uptake that *needs repair* (11 counts).

Table 33
Frequency and Distribution of Subcodes for Uptake of Recasts in Learner-System Interaction

Subcode	Definition	Example	Count	%
No uptake	The recast item does not occur anywhere in the learner's production after the recast move, neither in the learner's immediate response move nor in the further course of the interaction.	<p>Learner: yes, I will book <u>the</u> [error] alpha on the 12 October 2019</p> <p>System: Would you like me to book <u>room Alpha</u> [recast form] for the 12 October 2019? [recast]</p> <p>Learner: yes</p> <p>System: Thank you, the room has been booked for you.</p> <p>Could I help you with anything else?</p> <p>Learner: no, thank you very much and have a nice day. See you soon</p> <p>System: We are looking forward to welcoming you at One King Hotel. Goodbye.</p> <p>[End of interaction] [No uptake]</p>	20	50.0
Immediate uptake			4	10.0
Repair	The response move (i.e., the learner's turn immediately following the recast move) contains the target-like form or forms.	<p>Learner: Okey <u>the</u> [error] room Alpha would fit our expectations, is the room free on 12th October 2019?</p> <p>System: So, I will book [recast form] <u>room Alpha</u> on the 12th October 2019 for you. Could you please confirm if you would like me to proceed?</p> <p>Learner: I confirm the booking of <u>room Alpha</u> [Immediate uptake (Repair)], thanks</p>	3	

Subcode	Definition	Example	Count	%
Needs repair	The response move (i.e., the learner's turn immediately following the recast move) contains the recast item or items with the original error or a new error.	Learner: Yes sure. I'm talking about <u>the</u> [error] room Gamma System: <u>Room Gamma</u> [recast form] has two large screens and loudspeakers. The screens can be moved around. Learner: Thanks and how is the atmosphere in <u>the</u> [Immediate uptake (Needs repair)] Room Gamma?	1	
Delayed uptake			16	40.0
Repair	The learner produces the target-like form or forms later in the interaction rather than in the learner's turn immediately following the recast move.	[Message 23] Learner: But I will book a room, <u>wich</u> [error] room do you think is the best for my options? [Messages 24 - 25] System: Yes, sure. Could you please tell me <u>which</u> [target-like form] room you are referring to? [Messages 26 - 27] [...] [Message 28] Learner: can you tell me, <u>which</u> [Delayed uptake (Repair)] room is the best for my options?	5	
Needs repair	The learner produces the recast item or items with the original error or a new error in a turn later in the interaction, i.e., after the turn immediately following the recast move.	[Message 3] Learner: How many <u>person</u> [error] fit in there? [Message 4] System: There is space for 200 <u>people</u> [recast form] in room Gamma. [Messages 5 - 16] [...] [Message 17] Learner: Okay we are about 150 <u>person</u> [Delayed uptake (Needs repair)] is that okay? It's on the 12.10.2019	11	

In addition, instances of re-uptake were coded where the learner produced the recast item again after immediate or delayed uptake. Re-uptake may involve the use of the target-like form (*repair*) or the use of an erroneous form (*needs repair*). Examples for each subtype may be found in Table 34.

In the data, repaired re-uptake may be preceded by repaired uptake. In other words, learners may use the recast form correctly twice following a recast. In other instances, however, repaired re-uptake occurred after unrepaired uptake, as illustrated in the example for *re-uptake/repair* in Table 34. This means that learners used the recast form incorrectly in a first instance and correctly in the second instance following the recast. Similarly, unrepaired re-uptake may occur after either repaired or unrepaired uptake. So, learners may produce the target-like form in the first uptake but not the second (see the example for *re-uptake/needs repair* in Table 34) or they produced erroneous forms in both instances.

As the counts in Table 34 indicate, of the total 20 uptakes, 11 (55.0%) led to re-uptake. Of these, three were repaired (27.3%) and eight were unrepaired (72.7%).

Table 34

Frequency and Distribution of Subcodes for Re-Uptake of Recasts in Learner-System Interaction

Subcode	Definition	Example	Count	%
Re-uptake				
Repair	The next time the recast item or items occurs in the learner production after immediate or delayed uptake, it is/they are target-like.	<p>[Message 13] Learner: Both. Do they have a <u>beamer</u> [error] in?</p> <p>[Messages 14 – 16] System: Alpha is equipped with a state-of-the-art multimedia set including a big screen, <u>projector</u> [recast form] and sound system.</p> <p>Room Beta can be equipped with a moveable screen, <u>projector</u> [recast form], and loudspeakers.</p> <p>Room Gamma has two moveable monitors and speakers.</p> <p>[Messages 17 – 34] [...]</p> <p>[Message 35] Learner: and does gamma have a <u>beamer</u>?</p>	3	27.3

Subcode	Definition	Example	Count	%
		[Delayed uptake (Needs repair)]		
		[Message 36] System: The price for Gamma is 1100 CHF.		
		[Message 37] Learner: does the room have a <u>projector</u> ? ['Re-uptake' (Repair)]		
Needs repair	The next time the learner produces the recast item or items after immediate or delayed uptake, it contains / they contain the original error or a new error.	[Message 16] Learner: I think it would be <u>the</u> [error] room beta with additional 20 people [Message 17] System: Very well. Would you like to book <u>room Beta</u> [recast form]? [Messages 18 – 38] [Message 39] Learner: ok room beta with additional 20 people would be goog [Delayed uptake (Repair)] Messages 40 – 49: User: how much does <u>the</u> room beta coast? ['Re-uptake' (Needs repair)]	8	72.7

Recasts in SCMC

The following section presents the findings on recasts in the learner-peer SCMC interactions, highlighting the similarities and differences in the realization and frequency of each recast subcode in the two interaction types.

Table 35 presents an overview of the SCMC recast occurrences compared to those found in learner-system interaction. As can be seen, SCMC led to a total of 12 recasts overall, or 0.67 per interaction and 0.02 recasts per message. Compared to the occurrence of recasts in learner-system interaction, SCMC led to about three times fewer recasts overall as well as per interaction and half as many recasts per message.

Table 35*Frequency of Recasts in Learner-System Interaction and SCMC*

Measure	System	SCMC
Total recasts	40	12
Recasts per interaction	1.90	0.67
Indicators per message	0.04	0.02

Note. System: interactions: $n = 21$, messages: $n = 1139$. SCMC: interactions: $n = 18$, messages: $n = 611$.

Linguistic Focus

Table 36 shows a side-by-side overview of the linguistic focus of recasts found in SCMC and learner-system interaction. The examples are taken from the SCMC data. As this comparison shows, both SCMC and learner-system interaction showed a prevalence of recasts focusing on grammar. In SCMC, six or 54.6% of recasts addressed grammatical errors, compared to 26 or 65.0% in learner-system interaction. The second most frequent recast type found in SCMC was *spelling* with three counts or 27.3%, compared to a relatively lower frequency of six counts or 15.0% in learner-system interaction. In contrast to learner-system interaction where recasts of spelling and lexical errors were equally frequent with each six counts or 15.0% of all recasts, SCMC did not produce any recasts of lexical errors. However, SCMC produced a relatively higher occurrence of recasts focusing on multiple errors, with a total count of two multi-focus recasts of 18.2% of all recasts, compared to two counts or 5.0% in learner-system interaction.

Table 36

Frequency and Distribution of Subcodes for Linguistic Focus of Recasts in Learner-System Interaction and SCMC

Subcode	System		SCMC	
	Count	%	Count	%
Grammar	26 ^a	65.0	6	54.6
Spelling	6	15.0	0	0.0
Lexis	6	15.0	3	27.3
Multi-focus	2	5.0	2	18.2

Note. System: $n = 21$. SCMC: $n = 18$.

Grammar

Analyzing the linguistic focus of recasts in more detail, SCMC recasts in general addressed the same types of errors as the system recasts. For instance, in recasts focusing on grammatical errors, SCMC recasts equally addressed the omission of the indefinite article (Example 88: “nice impression”), the incorrect inflection for person of nouns (Example 87) and verbs (Example 88: “which one of these rooms are”), and the misselection of *an* instead of *a* (Example 89).

Example 87

Grammar recast on the incorrect inflection for person of nouns in SCMC

7	Learner	which one of this rooms takes place for 150 <u>peoples</u> ?	<i>Error</i>
8	Peer	I think the gamma room isthe best one, because it takes place for 200 <u>people</u>	<i>Recast form</i>
9	Peer	the beta room is too small, its for130 <u>people</u>	<i>Recast form</i>

Example 88

Grammar recast on the omission of the indefinite article and the incorrect inflection for person of verbs in SCMC

28	Learner	i ask again, which one of these rooms <u>are</u> the best for 150 people, a relax situation, so that we can work in groups, <u>nice impression</u> and a device for recording and audios. the cost sholud'nt be over 1500 CHF	<i>Error 1</i>
			<i>Error 2</i>
29	Peer	the best room <u>is</u> alpha room, 270 people its a relax situation, they can work in groups and <u>a nice impression</u> and device for recording and audios	<i>Recast form on error 1</i>
			<i>Recast form on error 2</i>
30	Peer	the price of the room <u>is</u> 1'400	<i>Repeated recast form 2 on error 1</i>

Example 89

Grammar recast on misselection of an instead of a in SCMC

4	Learner	I'm looking for a room where can fit at least 150 people. There need to be a projector with audio where we can project the ppp. We should have the possibility to talkin smaller groups and have <u>an</u> nice atmosphere. The room needs to impress the clients but not too much. The costs need to be under CHF 1'500.-	<i>Error</i>
5	Peer	In this case, we recommend you to take the room "Beta", it has place for 130 people inside and 60 outside. We offer you a projector and a big screen. The conversation in smaller groups is no problem. You will guaranteed have <u>a</u> nice time, the room ist on top floor so you can see the whole city, best atmosphere. The price is around CHF 900.-.	<i>Recast form</i>

However, the superfluous use of *the* with a room name which was the most frequently addressed error in system recasts was never addressed in SCMC. A grammatical error type – incorrect word order – was addressed in SCMC but not in learner-system interaction, as illustrated in Example 90. In this example, the learner

incorrectly places the adverb “also” after the main verb “do”. The correct order (“also take”) is provided in the recast move.

Example 90

Grammar recast on misselection of incorrect word order in SCMC

- | | | | |
|---|---------|--|--------------------|
| 7 | Learner | We <u>do also</u> a presentation and for that it is important for us to have a projector and to do conversations in small groups must be possible. Your offer for our event is very good! Wow i think the old decoration from 1920 has to be very impressiv and beautiful. | <i>Error</i> |
| 8 | Learner | The room must also have a good and relaxed atmosphere... | |
| 9 | Peer | You <u>could also</u> take a reservation for both rooms. You could do the presentation in the room Alpha, because there are fix chairs and later you could go to the room Gamme, where it would have enough place and space for the conversations. | <i>Recast form</i> |

Spelling

Recasts focusing on spelling in SCMC addressed the same types of spelling errors as did system interaction, i.e., the omission or misselection of letters as well as typing errors, as illustrated in

Example 91, where the learner’s spelling “p?rojektor” includes both a typing error as well as the misselection of “k” instead of “c” and the peer recasts the target-like “projector”.

Example 91

Spelling recast on the omission of letters and typing errors in SCMC

- | | | | |
|----|---------|--|--------------------|
| 11 | Learner | Has this room also a <u>p?rojektor</u> with audio | <i>Error</i> |
| 12 | Peer | The room must also have a good and relaxed atmosphere... | |
| 13 | Peer | no sorry | |
| 14 | Peer | we have a room with a <u>projector</u> and audio | <i>Recast form</i> |

Multi-focus

Of the two multi-focus recasts found in SCMC, one addressed grammar and spelling errors, while the other focused on grammar and lexis. The first multi-focus recast is presented in Example 92. As can be seen in the example, the recast move contains the target-like form of the misspelled words “Franks” and “regardes” as well as the target-like verb phrase construction with the adverb “also”, correctly placing it before the main verb.

Example 92*Multi-focus recast in SCMC*

1	Learner	Dear One King Hotel , here is Mergime from EJPD . I want need a room on the 12th okt. 2019 for a meeting. It should be for 150 people. I also need to show videos in there so it was good if there is a laptop and a projector. There <u>will be also</u> people that need to talk in smaller groups and they should be able to in this room. My budget is like 1500. <u>Franks</u> .	<i>Grammar error</i>
2	Learner	Kindest <u>regardes</u> Mergime	<i>Spelling error 2</i>
3	Peer	Dear Madame	
4	Peer	We have a room called Gamma where 200 people can fit. It has 2 big mobile screens. The price for that room is 1100 swiss francs. We <u>also have</u> a room called Alpha where 270 people can fit with the most modern screen it is even bigger than the one in Gamma. The price for that room is 1400 swiss <u>francs</u> . Kind <u>regards</u> Aida Colangelo Martinez	<i>Recast form on grammar error</i> <i>Recast forms on spelling error 1 and 2</i>

Example 93 presents the second multi-focus recast which provides the target-like plural “people” and the lexically correct phrase “have a presantation” (note incorrect spelling) as a recast for “have to presentation something”. It needs to be noted, however, that the recast form, while correcting the lexical error, introduces a new spelling error in the word “presantation”.

Example 93*Multi-focus recast in SCMC*

5	Learner	I need a room for 150 <u>peoples</u>	<i>Grammar error</i>
6	Learner	and i have to <u>presentation</u> somethin	<i>Lexis error</i>
7	Learner	something*	
8	Peer	I would recommend room Alpha	
9	Peer	Roomfor 270 <u>people</u>	<i>Recast form on grammar error</i>
10	Peer	And you can <u>have a presantation</u> there	<i>Recast form on lexis error</i>
11	Peer	It is like a Theater	

Intensity of Recast Within Interaction

With respect to the intensity of recasts, there was no repeated recast of the same form within the same interaction in SCMC, whereas in system interaction, singular (57.5%) and repeated (42.5%) incidences were almost balanced (see Table 37).

Table 37

Frequency and Distribution of Subcodes for Intensity of Recast Within Interaction in Learner-System Interaction and SCMC

Subcode	System		SCMC	
	Count	%	Count	%
Singular incidence	23	57.5	11	100.0
Repeated incidence	17	42.5	0	0.0

Intensity of Recast Form Within Recast Move

With regard to the repeated incidence of the recast form within the same recast move, however, SCMC showed a higher relative frequency of repeated incidence versus singular incidence. In SCMC recasts, the recast form was repeated in three or 27.3% of cases, in system recasts in six or 15.0% of cases (see Table 38).

Table 38

Frequency and Distribution of Subcodes for Intensity of Recast Form Within Recast in Learner-System Interaction and SCMC

Subcode	System		SCMC	
	Count	%	Count	%
Singular incidence	34	85.0	8	72.7
Repeated incidence	6	15.0	3	27.3

Completeness

As opposed to learner-system interaction where about half of the recast moves (45.0%) were *complete*, i.e., addressed all learner errors, while the other half (55.0%) were *partial*, i.e., left some of the learner errors unaddressed, all 11 SCMC recasts (100.0%) were partial recasts (see Table 39).

Table 39

Frequency and Distribution of Subcodes for Completeness of Recasts in Learner-System Interaction and SCMC

Subcode	System		SCMC	
	Count	%	Count	%
Complete	18	45.0	0	0.0
Partial	22	55.0	11	100.0

Length

With regard to their length, the majority of SCMC recast moves consisted of multiple message (eight counts or 72.7%), while the remaining moves consisted of multiple sentences (two counts or 18.2%) or a single sentence (one count or 9.1%). This pattern differs from the system recasts where the recast move most frequently (47.5% of instances), consisted of a single sentence and multi-message-moves were the most infrequent type (25.0% of instances). Frequencies and distribution of the subcodes for length of recasts in the two interaction types are displayed in Table 40.

Table 40

Frequency and Distribution of Subcodes for Length of Recast in Learner-System Interaction and SCMC

Subcode	System		SCMC	
	Count	%	Count	%
Single sentence	19	47.5	1	9.1
Multiple sentences	11	27.5	2	18.2
Multiple messages	10	25.0	8	72.7

Uptake

As opposed to learner-system interaction where 50.0% of the recast items were followed by learner uptake in the remainder of the interaction, there was relatively less uptake in SCMC interaction (see Table 41): Of the total 11 recasts, nine or 81.8% were not followed by uptake, compared to two or 18.2% recasts followed by uptake. One of the uptakes occurred immediately following the recast move, one was delayed.

In SCMC, the immediate uptake contained the unrepaired form, whereas the delayed uptake led to repair. This contrasts with learner-system interaction where immediate uptake more frequently led to repair than non-repair and also more frequently led to repair than delayed uptake. However, the low incidence of uptake in SCMC prevents a conclusive statement on the comparative prevalence and nature of repair in the two interaction types.

Table 41

Frequency and Distribution of Subcodes for Uptake of Recasts in Learner-System Interaction and SCMC

Subcode	System		SCMC	
	Count	%	Count	%
No uptake	20	50.0	9	81.8
Immediate uptake	4	10.0	1	9.1
Repair	3		0	
Needs repair	1		1	
Delayed uptake	16	40.0	1	9.1
Repair	5		1	
Needs repair	11		0	

As depicted in Table 42, no re-uptake of recast forms was observed in SCMC. In comparison, 11 out of 20 instances or 55.0% of uptake in system interaction were followed by re-uptake.

Table 42

Frequency and Distribution of Subcodes for Re-Uptake of Recasts in Learner-System Interaction and SCMC

Subcode	System		SCMC	
	Count	%	Count	%
No re-uptake	9	45.0	12	100.0
Re-uptake	11	55.0	0	0.0
Repair	3		0	
Needs repair	8		0	

Learners' Perception on Recasts

No additional data on the learners' perception on recasts were gained from the questionnaires and interviews. When presented with a recast sequence during interviews, participants from both groups did not comment on the recast.

4.2.3 Lexical Alignment

As described above (see section Coding of Alignment), lexical alignment was operationalized as the reuse of n-grams of three or more words within five messages (learner-system interaction) or ten messages (SCMC), respectively. As also described, all instances of lexical alignment in the learner-system scripts were coded and assigned a subcode specifying the direction of alignment (i.e., *learner to interlocutor*, *learner to learner (self)*, etc.). The following paragraphs present the findings of this analysis.

Lexical Alignment in Learner-System Interaction

Overview

Table 43 shows an overview of the number of occurrences for each direction of lexical alignment found in learner-system interaction. For better illustration and

comparability, the ratio of lexical alignment per interaction (occurrences/total number of interactions) and per message (occurrences/total number of messages) are also listed.

As shown in Table 43, learners aligned to their own or the interlocutor's production in a total of 87 occurrences (4.143 per interaction). The most frequent type of learner alignment was alignment from learner to learner (self) with a total of 62 occurrences (2.952 per interaction). In a total of 19 occurrences (0.905 per interaction), the learner aligned to the interlocutor. In the remaining six occurrences (0.286 per interaction), the learner aligned to both, i.e., reused an n-gram that occurred in their own as well as their interlocutor's production within the five preceding messages. There was a total of 87 occurrences (4.143 per interaction) of unprimed learner reuse, i.e., the reuse of an item outside the five-message-threshold in the learner production.

Focusing on interlocutor alignment, alignment from interlocutor to learner amounted to a total of 39 occurrences (1.857 per interaction). It needs to be noted that in almost half of the cases (16 occurrences), the interlocutor reused an n-gram containing a learner error. An additional 12 occurrences (0.571 per interaction) were unprimed interlocutor reuse, where the interlocutor reused an n-gram from the learner's production outside the threshold of five messages. As mentioned above (methods), instances of alignment from interlocutor to interlocutor were not considered for lack of interpretative value.

Table 43*Frequency of Subcodes for Lexical Alignment in Learner-System Interaction*

Subcode	Count	Per interaction (<i>n</i> = 21)	Per message (<i>n</i> = 1139)
Used by learner	174	8.286	0.153
Learner to learner	62	2.952	0.054
Learner to interlocutor	19	0.905	0.017
Learner to both	6	0.286	0.005
Total learner alignment	87	4.143	0.076
Learner unprimed ^a	87	4.143	0.076
Used by interlocutor	51	2.429	0.045
Interlocutor to learner	39 ^b	1.857	0.034
Interlocutor unprimed ^a	12	0.571	0.011

^a Unprimed counts exclude the first occurrences of n-grams in an interaction, although first occurrences were included in the unprimed code during coding for coding consistency reasons. To obtain the unprimed counts excluding first occurrences, all first occurrences were deducted by hand from the unprimed count in MAXQDA. ^b Of the 39 instances of interlocutor-to-learner alignment, 16 instances involved the repetition of a learner error.

Qualitative Findings

The following presents a more detailed, qualitative account of the lexical alignment found in learner-system interaction. The descriptions are presented separately for each direction of alignment.

As a general observation, lexical alignment often involved the reuse of items that consisted of more than three words and up to the length of an entire clause with seven or eight words, as will become evident in the examples below.

Learner to Learner (Self)

As the examples below show, lexical self-alignment by the learner occurred in a number of different contexts. In a first type of context, the learner reused an n-gram as part of a request that was structurally identical to the request that contained the initial occurrence of the item, while changing its reference, typically to a different room or requirement.

Example 94 and Example 95 illustrate this type of context. In Example 94, the learner initially asks about the “atmosphere” in room Alpha in message 28, followed by a question referring to the “atmosphere” in room Gamma. As can be seen, the learner reuses the 7-gram “and what about the atmosphere in room” from the previous question on the “atmosphere” referring to room Alpha. Similarly, in Example 95, the learner states that the room should offer the possibility to “make a presentation” in message 9 using the 5-gram “must be possible to make”. The same item is used again in line 11 to state a different requirement, i.e., “make little groups to share talks”.

Example 94

Lexical alignment from learner to learner (self) involving a change of reference in learner-system interaction

28	Learner	<u>And what about the atmosphere in room Alpha?</u>	<i>Learner unprimed</i>
29	System	Room Alpha is a spacious room located on the basement floor of the hotel. It is an auditorium furnished with a stage, rows of seats, and a multimedia and lighting system.	
30	Learner	What are the qualities of the Gamma room?	
31	System	Room Gamma would be CHF 1100.	
32	Learner	<u>And what about the atmosphere in room Gamma?</u>	<i>Learner to learner (self)</i>

Example 95

Lexical alignment from learner to learner (self) involving a change of reference in learner-system interaction

9	Learner	it <u>must be possible to make</u> a presentation	<i>Learner unprimed</i>
10	System	Room Gamma has two moveable monitors and speakers.	
11	Learner	it's also <u>must be possible to make</u> little groups to share talks	<i>Learner to learner (self)</i>

As the following examples demonstrate, the learners' lexical self-alignment may also occur as part of negotiation of meaning. More precisely, learners may reuse an n-

gram from an utterance that caused non-understanding during the response stage of the negotiation sequence. This occurred in responses of the subtypes *elaboration*, *reduction*, *modification*, and naturally, *repetition*. Example 96 shows an instance of the response type *elaboration*. As can be seen in the example, the learner in message 71 reuses the 4-gram “I search a room” from message 69.

Example 96

Lexical alignment from learner to learner (self) as part of negotiation of meaning
response type Elaboration in learner-system interaction

69	Learner	<u>I search a room</u>	<i>Learner unprimed</i>
70	System	I am afraid I have trouble understanding. Please could you rephrase?	
71	Learner	<u>i search a room</u> with place for 150 people	<i>Learner to learner (self)</i>

In Example 97, the learner’s *reduction* response in message 10 contains the same 6-gram “space for at least 150 peoples” as the trigger utterance.

Example 97

Lexical alignment from learner to learner (self) as part of negotiation of meaning
response type Reduction in learner-system interaction

8	Learner	The should have <u>space for at least 150 peoples</u> , have a projector for a presentation, allow conversations in small groups, have a good atmosphere, should look quite noble but not over the top and shouldn't cost more than 1500 Francs.	<i>Learner unprimed</i>
9	System	So sorry, I could not understand your last message. Could you please say it again?	
10	Learner	Have you got any rooms with <u>space for at least 150 Peoples</u> ?	<i>Learner to learner (self)</i>

Example 98 illustrates a response of the subtype *modification* that reuses two n-grams from the trigger utterance “I would like to”, and “a room for 1500 people”.

Example 98

Lexical alignment from learner to learner (self) as part of negotiation of meaning

response type Modification in learner-system interaction

- | | | | |
|----|---------|---|----------------------------------|
| 11 | Learner | I would like to book <u>a room for 1500</u>
<u>peoples</u> . | <i>Learner unprimed</i> |
| 12 | System | Excuse me, can you please rephrase your last message? | |
| 13 | Learner | I would like to reserve <u>a room for 1500</u>
<u>people</u> please. | <i>Learner to learner (self)</i> |

In Example 99 and Example 100, the response involves a repetition of the trigger utterance and thus complete lexical overlap between the two utterances.

Example 99

Lexical alignment from learner to learner (self) as part of negotiation of meaning

response type Repetition in learner-system interaction

- | | | | |
|----|---------|---|----------------------------------|
| 9 | Learner | I would like to book <u>a room for 1500</u>
<u>peoples</u> . | <i>Learner unprimed</i> |
| 10 | System | Excuse me, can you please rephrase your last message? | |
| 11 | Learner | I would like to reserve <u>a room for 1500</u>
<u>people</u> please. | <i>Learner to learner (self)</i> |

Example 100

Lexical alignment from learner to learner (self) as part of negotiation of meaning

response type Repetition in learner-system interaction

- | | | | |
|----|---------|--|----------------------------------|
| 9 | System | The price for Gamma is 1100 CHF. | |
| 10 | Learner | <u>and the others?</u> | <i>Learner unprimed</i> |
| 11 | System | Room Gamma has two moveable monitors and speakers. | |
| 12 | Learner | <u>and the others?</u> | <i>Learner to learner (self)</i> |

Learner to Interlocutor

Lexical alignment from learner to interlocutor in learner-system interaction typically involved adjacency pairs such as question-answer or thanking and leave-taking exchanges. The following examples illustrate question-answer pairs where the interlocutor requests some information from the learner and the learner in their reply reused an n-gram from the interlocutor's question turn.

In Example 101, the learner's answer reuses the noun phrase "the number of guests" from the interlocutor's question.

Example 101

Lexical alignment from learner to interlocutor in question-answer pairs in learner-system interaction

44	System	Please could you give me <u>the number of guests</u> ?	<i>System unprimed</i>
45	Learner	I would like to book the room Alpha for 12 october 2019 and <u>the number of guests</u> is minimum 150 persons	<i>Learner to interlocutor</i>

In Example 102, the overlapping item is the verb phrase "would be" and the definite article "the" of the subsequent noun phrase stating the booking date.

Example 102

Lexical alignment from learner to interlocutor in question-answer pairs in learner-system interaction

30	System	What <u>would be the</u> date for the booking?	<i>System unprimed</i>
31	Learner	That <u>would be the</u> 12 of October 2019.	<i>Learner to interlocutor</i>

Similarly, alignment from learner to interlocutor frequently occurred in thanking and leave-taking sequences. In Example 103, the learner's response to the interlocutor's thanking and leave-taking turn shows complete lexical overlap with the interlocutor's utterance.

Example 103

Lexical alignment from learner to interlocutor in thanking and leave-taking sequences in learner-system interaction

39	System	<u>Thank you very much. Goodbye.</u>	<i>System unprimed</i>
40	Learner	<u>Thank you very much. Goodbye.</u>	<i>Learner to interlocutor</i>

In Example 104, the learner reuses the item “have a nice day” from the interlocutor’s preceding turn.

Example 104

Lexical alignment from learner to interlocutor in thanking and leave-taking sequences in learner-system interaction

38	System	Thank you for your reservation and <u>have a nice day.</u>	<i>System unprimed</i>
39	Learner	<u>have a nice day.</u>	<i>Learner to interlocutor</i>

On two occasions of alignment from learner to interlocutor, however, the learners’ alignment was not part of an adjacency pair. Rather, these sequences involved, similarly to one of the contexts seen in learner self-alignment above, a shift in reference while maintaining the syntactic structure between the interlocutor’s and the learner’s utterances.

In the first instance, presented in Example 105, the learner similarly reuses the interlocutor’s item “space for up [to]” from message 19 to form a request on the same requirement, i.e., room size, in message 20.

Example 105

Lexical alignment from learner to interlocutor involving a change of reference in learner-system interaction

19	System	Room Alpha has <u>space for up</u> to 270 people.	<i>System unprimed</i>
20	Learner	Look.. i need a Room wih <u>space for up</u> 150 People	<i>Learner to interlocutor</i>

In the other instance, presented in Example 106, the dialog system's turn in message 11 contains information on the size of room Beta. This is followed by the learner in message 12 requesting information on the technical equipment of room Beta, thus shifting to a different requirement for the same room as in message 11. At the same time, the learner in message 12 performs a shift between the dialog system's turn in message 9 on the equipment ("monitors and speakers") available in room Gamma to the equipment in room Beta. As can be seen, the learner accordingly reuses two items from the interlocutor's preceding production, "the Beta room" from message 11 and "monitors and speakers" from message 9.

Example 106

Lexical alignment from learner to interlocutor involving a change of reference in learner-system interaction

9	System	Room Gamma has two moveable <u>monitors and speakers</u> .	<i>System unprimed 1</i>
10	Learner	How much space has it? Can 150 people fit in there?	
11	System	<u>The Beta room</u> has space for a total of 190 people, indoor and outdoor.	<i>System unprimed 2</i>
12	Learner	Has <u>the Beta room</u> also got <u>monitors and speakers</u> ?	<i>Learner to interlocutor 1 and 2</i>

Additionally, in one instance, the learner reused an item from the interlocutor's production in the context of a meaning negotiation, as also found in learner self-alignment (see above). The extract is presented in Example 107. As can be seen, the learner's *modification* response in message 28 integrates the lexical item "tell me which room" from the interlocutor's utterance in message 25.

Example 107

Lexical alignment from learner to interlocutor in negotiation of meaning in learner-system interaction

25	System	Could you please <u>tell me which room</u> you are referring to?	<i>System unprimed</i>
----	--------	--	------------------------

- | | | | |
|----|---------|--|--|
| 19 | System | Please could you confirm if you <u>would like to book room Alpha for the 12th of October 2019?</u> | <i>Interlocutor to learner 2</i> |
| 20 | Learner | I hereby confirm, that I <u>would like to book room Alpha for thw 12th of October 2019.</u> | <i>Learner to both 2 / Learner to learner (self) 1</i> |

Interlocutor to Learner

The majority of alignment from interlocutor to learner involved the confirmation of a booking request. The learner in these instances stated a booking request which was followed by the interlocutor's request to confirm. In Example 109, for instance, the interlocutor's request to confirm reuses the booking date, often along with an additional element such as the room name (here: "ro[o]m Gamma") or the terms "book" or "reserve" from the learner's booking request.

Example 109

Lexical alignment from interlocutor to learner involving the confirmation of a booking request in learner-system interaction

- | | | | |
|----|---------|--|--------------------------------|
| 62 | Learner | <u>romm Gamma for the 12 october 2019</u> | <i>Learner unprimed</i> |
| 63 | System | Would you like me to book <u>room Gamma for the 12 october 2019?</u> | <i>Interlocutor to learner</i> |

As Example 109 shows, alignment from interlocutor to learner may also involve spelling correction. As seen in the example, since misspellings were disregarded, i.e., previously corrected, for the analysis of lexical alignment, the interlocutor's "room Gamma" in message 63 was considered a reuse of the learner's misspelled "romm Gamma" in message 62. As a note, these instances of lexical alignment are at the same time part of a recast sequence. In the above extract, for example, the noun phrase "room Gamma" in message 63 embedded in the reused 8-gram "room Gamma for the 12 october 2019" is at the same time a recast of "romm Gamma" in message 62.

Conversely, alignment from interlocutor to learner may involve the reuse of an erroneous learner item. This phenomenon was limited to the reuse of erroneous booking dates. The reuse of misspelled booking dates was due to technological

reasons, as the dialog system was designed to reproduce the date item as sent by the learner. The reuse of a misspelled booking date is illustrated in Example 110. As can be seen, the learner's spelling ("oktober") as well as grammatical errors (use of cardinal instead of ordinal number; omission of "of") from line 52 are reproduced by the interlocutor in line 53.

Example 110

Lexical alignment from interlocutor to learner involving the reuse of an erroneous booking date in learner-system interaction

52	Learner	the date is on <u>the 12 oktober</u>	<i>Learner unprimed</i>
53	System	I will reserve room Gamma for <u>the 12 oktober</u> . Could you please confirm if this is correct?	<i>Interlocutor to learner</i>

In a few instances of alignment from interlocutor to learner, the interlocutor reproduced an item from the learner's leave-taking or thanking-expressions such as "thank you so much" or "have a nice day", as in Example 111.

Example 111

Lexical alignment from interlocutor to learner in a leave-taking sequence in learner-system interaction

25	Learner	No, that would be everything. Thank you for your help and <u>have a nice day!</u>	<i>Learner unprimed</i>
26	System	Thank you very much for your reservation. <u>Have a nice day.</u>	<i>Interlocutor to learner</i>

Learner Unprimed

The following subsection analyzes the learners' unprimed reuse of lexical material, i.e., the reuse of n-grams beyond the threshold. As mentioned above, the first occurrence of a considered n-gram in a given interaction was also assigned the code *unprimed* for coding consistency. However, first occurrences are not included in the following analysis.

With respect to the learner's unprimed use of n-grams, it was found that it occurred in the same contexts as lexical alignment occurring within the five-message-threshold as presented above. Thus, unprimed use of lexical material may involve a shift of reference, as illustrated in the following examples. In Example 112, the learner in message 1 enquires about the possibility "to play video and audio". In message 12, the same inquiry refers exclusively to room Alpha.

Example 112

Unprimed learner reuse of n-grams involving change of reference in learner-system interaction

1	Learner	Hello, i am searching for a room suitable for our company event that costs 1'500 Francs maximum. It will take place on 12th October 2019. The room should be big enough to fit a minimum of 150 people. We also need to <u>be able to play video and audio</u> in the room.	<i>Learner unprimed</i>
2 – 11		[...]	
12	Learner	Will we <u>be able to play video and audio in room alpha</u> ?	<i>Learner unprimed</i>

Similarly, in Example 113, the learner reuses the item "it must be" from message 9 again six messages down in message 15, shifting from the requirement "to make a presentation" to the requirement of having "a good vibe". Note that the extract also contains separate alignment from learner to learner (self) in message 11, where the learner reuses the 5-gram "must be possible to make" from message 9.

Example 113

Unprimed learner reuse involving a change of reference in learner-system interaction

9	Learner	<u>it must be possible to make</u> a presentation	<i>Learner unprimed 1 (First occurrence) / Learner unprimed 2 (First occurrence)</i>
10	System	Room Gamma has two moveable monitors and speakers.	

11	Learner	it's also <u>must be possible to make</u> little groups to share talks	<i>Learner to learner (self) 1</i>
12 – 14		[...]	
15	Learner	<u>it must be</u> a good vibe in the room for the people	<i>Learner unprimed 2</i>

Furthermore, unprimed use frequently occurred in the context of rephrasing, as it did within-threshold learner alignment. The following examples illustrate this pattern. In Example 114, the learner asks about the requirement “atmosphere” in message 14. After the interlocutor responds inappropriately in the subsequent turn, the learner first shifts to a different requirement “make also private speaks” before coming back to the “atmosphere” in message 23. As can be seen, the learner in their reformulation in message 23 reuses “a good atmosphere” that first occurred nine messages previously. Note that the utterance in message 23 additionally contains the learner self-alignment “in which room can i” reused within the five-message threshold.

Example 114

Unprimed learner reuse involving rephrasing in learner-system interaction

14	Learner	okey which room is the best for <u>a good atmosphere</u>	<i>Learner unprimed 1 (First occurrence)</i>
15	System	Yes, sure.	
16	System	Yes, certainly.	
17	Learner	<u>In which room can I</u> make also private speaks	<i>Learner unprimed 2 (First occurrence)</i>
18 – 22		[...]	
23	Learner	<u>in which room can i</u> have <u>a good atmosphere?</u>	<i>Learner unprimed 1 and 2</i>

Similarly, Example 115 shows how a learner initially uses the items “in your hotel” and “for minimum 150 people” in messages 4 and 6, respectively, and again in message 18. As becomes evident, the learner’s turn in message 18 can be understood as a ‘restart’ following the interlocutor’s request to “please try one more time” in message 17.

Example 115*Unprimed learner reuse involving rephrasing in learner-system interaction*

4	Learner	I am searching for a room <u>in your hotel</u>	<i>Learner unprimed 1 (First occurrence)</i>
5	System	We have multiple rooms available.	
6	Learner	So the room i'm searching for should be big enough <u>for minimum 150 people</u>	<i>Learner unprimed 2 (First occurrence)</i>
7 – 16		[...]	
17	System	I am sorry, I did not understand that. Could you please try one more time?	
18	Learner	Yes sure, I would like to book a room <u>for minimum 150 people in your hotel</u> . The room should let us do a presentation and also meeting for small groups. It should also have a good atmosphere. It should make the guests happy. Do you have room for maximum 1500 Franks?	<i>Learner unprimed 1 and 2</i>

Finally, unprimed learner use of n-grams may involve the reuse of thanking formulas. In within-threshold alignment, aligned phatic discourse was typically part of adjacency pairs, such as thanking or leave-taking routines. As can be seen in the following example, learners may also reuse the same thanking-formulas outside the threshold.

Example 116*Unprimed learner reuse as part of a thanking sequence in learner-system interaction*

13	Learner	<u>Thank you for</u> contacting One King Hotel. I will gladly assist you in finding a room.	<i>System unprimed (First occurrence)</i>
14 – 24		[...]	
25	Learner	No, that would be everything. <u>Thank you for</u> your help and have a nice day!	<i>Learner unprimed</i>

Interlocutor Unprimed

Focusing on the interlocutor's unprimed reuse of lexical material, it was found to include a slightly more diverse set of lexical items than within-threshold alignment from interlocutor to learner which, as shown above, almost exclusively occurred in the

context of booking date confirmations. Unprimed interlocutor use of n-grams typically occurred, on the one hand, around booking dates and room names (see Example 117) as well as leave-taking and thanking (see Example 118 and Example 119). In these instances, the reuse of n-grams was delayed to beyond five messages of their first use, typically because there was initial non-understanding and the learner had to restate their request.

Example 117

Unprimed interlocutor reuse as part of a booking confirmation in learner-system interaction

62	Learner	<u>romm Gamma for the 12 October 2019</u>	<i>Learner unprimed (First occurrence)</i>
63 – 68		[...]	
69	Learner	I will reserve <u>room Gamma for the 12 October 2019</u> . Could you please confirm if this is correct?	<i>Interlocutor unprimed</i>

Example 118

Unprimed interlocutor reuse as part of a leave-taking and thanking sequence in learner-system interaction

60	Learner	so <u>goodby and thank you for</u> the reservation	<i>Learner unprimed (First occurrence)</i>
61 – 73		[...]	
74	System	I wish you a very nice day. <u>Goodbye and thank you for</u> choosing One King Hotel.	<i>Interlocutor unprimed</i>

Example 119

Unprimed interlocutor reuse as part of a leave-taking and thanking sequence in learner-system interaction

58	Learner	<u>Thank you very much</u>	<i>Learner unprimed (First occurrence)</i>
59 – 72		[...]	
73	System	Thank you so much.	
74	Learner	Thank you	

75	System	<u>Thank you very much</u> for choosing One King Hotel.	<i>Interlocutor unprimed</i>
----	--------	---	------------------------------

On the other hand, unprimed interlocutor reuse of n-grams involved vocabulary pertinent to the task such as for instance “the cost for” (Example 120) or verb phrases used to form questions such as “can you please” (Example 121) or “could you give me” (Example 122). The examples also show that unprimed interlocutor reuse may occur within relatively close proximity to the first use, as in Example 121 (within seven messages) as well as at a far distance, as in Example 120 (within 35 messages) and Example 122 (within 24 messages).

Example 120

Unprimed interlocutor reuse of task-related vocabulary in learner-system interaction

28	Learner	Yes how much is <u>the costen for</u> this room?	<i>Learner unprimed (First occurrence)</i>
29 – 62		[...]	
63	System	<u>The cost for</u> Beta would be CHF 900. The terrace is included in the price.	<i>Interlocutor unprimed</i>

Example 121

Unprimed interlocutor reuse of task-related vocabulary in learner-system interaction

3	Learner	<u>Can you please</u> reed my message again ? It's fort he 12 Oktober 2019	<i>Learner unprimed (First occurrence)</i>
4 – 9		[...]	
10	System	Excuse me, <u>can you please</u> rephrase your last message?	<i>Interlocutor unprimed</i>

Example 122

Unprimed interlocutor reuse of task-related vocabulary in learner-system interaction

3	Learner	I'm not sure wich room would be the best option, <u>could you give me</u> a few informations?	<i>Learner unprimed (First occurrence)</i>
4 – 26		[...]	
27	System	Please <u>could you give me</u> the number of guests?	<i>Interlocutor unprimed</i>

Lexical Alignment in SCMC

The following section presents the findings on lexical alignment in SCMC, highlighting the parallels and differences in lexical alignment between the two interaction types.

Overview

The overview in Table 44 shows the occurrences of lexical alignment found in the two interaction types side by side. To correct for unequal numbers of interactions and messages in the two interaction types, the counts per interaction and per message are also stated.

As can be seen in the overview, uncorrected for number of messages, learner-system interaction produced almost three times as much total lexical reuse (i.e., within-threshold alignment and unprimed reuse across all directions) as SCMC (225 system vs. 78 SCMC). Looking at the individual alignment directions, learner-system interaction also produced about twice as much within-threshold learner alignment as SCMC (87 system vs. 41 SCMC). Similarly, unprimed learner reuse was about ten times higher in system interaction than in SCMC (87 system vs. eight SCMC). Overall, learner-system interaction produced about 3.5 times more learner reuse of lexical material (i.e., within-threshold learner alignment and unprimed learner reuse) than SCMC (174 system vs. 49 SCMC). Similarly, within-threshold alignment from interlocutor to learner in learner-system interaction was about 1.5 times higher than in SCMC (39 system vs. 25 SCMC). Unprimed interlocutor reuse was three times higher in learner-system interaction than in SCMC (12 system vs. four SCMC).

Corrected for number of messages, total lexical reuse across all directions was still about 1.5 times higher in learner-system interaction compared to SCMC (0.198 system vs. 0.128 SCMC). However, with respect to learner reuse, the frequency of within-threshold learner alignment per message was only 1.1 times higher in learner-system interaction compared to SCMC (0.076 system vs. 0.067 SCMC). With respect to the individual directions of within-threshold learner alignment, there was 1.2 times more alignment per message from learner to learner (self) in learner-system interaction (0.054 per message) compared to SCMC (0.043 per message). Conversely,

alignment per message from learner to interlocutor was 1.2 times lower in learner-system interaction (0.017 per message) than in SCMC (0.021 per message). Alignment per message from learner to both was 1.7 times higher in learner-system (0.005 per message) than in SCMC (0.003 per message). Unprimed learner reuse per message was eight times higher in learner-system interaction (0.076 per message) than in SCMC (0.013 per message). Overall, when corrected for number of messages, total learner reuse (i.e., within-threshold learner alignment and unprimed learner reuse) was almost twice as high in learner-system interaction (0.153 per message) compared to SCMC (0.080 per message). With respect to interlocutor alignment when corrected for number of messages, alignment from interlocutor to learner was 1.3 times lower in learner-system interaction (0.034 per message) than in SCMC (0.041 per message). Unprimed interlocutor reuse per message was about 1.5 times higher in learner-system interaction (0.011 per message) than SCMC (0.007 per message).

Comparing the distributions of the different alignment directions in the two interaction types, in learner-system interaction, the proportions of within-threshold learner alignment and unprimed learner reuse were exactly equal (87 within-threshold vs. 87 unprimed). In SCMC, the proportion of within-threshold learner alignment was considerably higher compared to unprimed learner reuse (41 within-threshold vs. eight unprimed). As a parallel between the two interaction types, alignment from learner to learner (self) was the most frequent direction of learner alignment, alignment from learner to interlocutor was the second most frequent, and alignment from learner to both was the least frequent. However, in learner-system interaction there was a clearer prevalence of learner self-alignment compared to the other directions of learner alignment, particularly compared to alignment from learner to interlocutor: In learner-system interaction, learner self-alignment (62 occurrences) was about three times higher than alignment from learner to interlocutor (19 occurrences) and ten times higher than alignment from learner to both (six occurrences). In SCMC, learner self-alignment (26 occurrences) was only twice as high as alignment from learner to interlocutor (13 occurrences) and 13 times higher than alignment from learner to both (two occurrences). With respect to interlocutor alignment, in both interaction types, within-threshold alignment from interlocutor to learner was higher compared to unprimed interlocutor reuse (system: 39 within-

threshold vs. 12 unprimed; SCMC: 25 within-threshold vs. four unprimed). This means that, in learner-system interaction, within-threshold interlocutor alignment was about three times higher than unprimed interlocutor reuse, while in SCMC within-threshold interlocutor alignment was about six times higher than unprimed interlocutor reuse.

Table 44

Frequency of Lexical Alignment in Learner-System Interaction and SCMC

Subcode	System			SCMC		
	Total count	Per interaction (<i>n</i> = 21)	Per message (<i>n</i> = 1139)	Total count	Per interaction (<i>n</i> = 18)	Per message (<i>n</i> = 611)
Reused by learner	174	8.286	0.153	49	2.722	0.080
Learner to learner	62	2.952	0.054	26	1.444	0.043
Learner to interlocutor	19	0.905	0.017	13	0.722	0.021
Learner to both	6	0.286	0.005	2	0.111	0.003
Total learner alignment	87	4.143	0.076	41	2.278	0.067
Learner unprimed ^a	87	4.143	0.076	8	0.444	0.013
Reused by interlocutor	51	2.429	0.045	29	1.611	0.047
Interlocutor to learner	39 ^b	1.857	0.034	25 ^c	1.389	0.041
Interlocutor unprimed ^a	12	0.571	0.011	4	0.222	0.007
Total	225	10.714	0.198	78	4.333	0.128

^a *Unprimed* counts exclude the first occurrences of n-grams in an interaction, although first occurrences were included in the *unprimed* code during coding for coding consistency reasons. To obtain the *unprimed* counts excluding first occurrences, all first occurrences were deducted by hand from the *unprimed* count in MAXQDA. ^b Of the 39 instances of interlocutor-to-learner alignment in learner-system interaction, 16 instances involved the repetition of a learner error. ^c Of the 25 instances of interlocutor-to-learner alignment in SCMC, nine instances involved the repetition of a learner error.

Qualitative Findings

Analyzing alignment in SCMC qualitatively and comparing it with the findings from learner-system interaction revealed the following observations.

Learner to Learner (Self)

Learner self-alignment within the given threshold of ten messages in SCMC could be observed in various contexts. As shown above, learner self-alignment in learner-system interaction primarily occurred in the contexts of reference shifts and negotiation of meaning. Lexical alignment in the context of reference shifts was also common in SCMC interaction. This context is illustrated in Example 123 and Example 124. In Example 123, the learner enquires about the cost of room Alpha in message 11 using the item “And how much cost the room”. In message 16, the learner reuses the n-gram to ask about the price of room Gamma.

Example 123

Lexical alignment from learner to learner (self) involving a change of reference in SCMC

11	Learner	<u>And how much cost the room</u> Alpha?	<i>Learner unprimed (First occurrence)</i>
12 – 15		[...]	
16	Learner	okay <u>and how much cost the room</u> Gamma	<i>Learner to learner</i>

Similarly, in Example 124, the learner uses the n-gram “The room must” to state the size requirement in message 3. The learner reuses the same n-gram in message 8 where they mention the “atmosphere” requirement.

Example 124

Lexical alignment from learner to learner (self) involving a change of reference in SCMC

3	Learner	<u>The room must</u> be big enough for 150 people and the price limit is about 1'500 francs. Is that possible?	<i>Learner unprimed (First occurrence)</i>
4 – 7		[...]	
8	Learner	<u>The room must</u> have a good and relaxed atmosphere...	<i>Learner to learner</i>

However, in SCMC, no lexical learner self-alignment was found in the context of negotiation of meaning. Rather, learners in SCMC frequently reused lexical material when specifying, correcting, or restating a previous utterance without being prompted to do so through an indication of non-understanding. Example 125 shows an instance where the learner restates their booking intentions and the booking date.

Example 125

Lexical alignment from learner to learner (self) in a restatement in SCMC

5	Learner	That sounds great. So is the room available <u>on the 12. October 2019</u> ?	<i>Learner unprimed (First occurrence)</i>
6	Peer	Yes, the room is available the whole day.	
7	Learner	Alright, then I'd like to reserve the room "Beta" <u>on the 12. October 2019</u> for the whole day.	<i>Learner to learner</i>

Similarly, in Example 126, the learner restates their booking intentions in message 31 after the initial booking request in message 26 using the same phrase “book this room”.

Example 126

Lexical alignment from learner to learner (self) in a restatement in SCMC

26	Learner	i need to <u>book this room</u> for the 12th of october 2019	<i>Learner unprimed (First occurrence)</i>
27 – 30		[...]	
31	Learner	that's very good can i <u>book this room</u> now?	<i>Learner to learner</i>

In Example 127, the learner rephrases the requirement from message 12 that the room should not be too ‘luxurious’ in message 16 as “It doesn’t have to be very big”, in both instances using the n-gram “it doesn’t have to be”.

Example 127

Lexical alignment from learner to learner (self) in a restatement in SCMC

12	Learner	I hope the room is nice decorated but <u>it doesn't have to be</u> too much luxury	<i>Learner unprimed (First occurrence)</i>
9 – 13		[...]	
16	Learner	<u>It doesn't have to be</u> very big, it should give our inviters the oppurtunity to talk in small groups and it should show some good atmosphary	<i>Learner to learner</i>

Learner to Interlocutor

Alignment from learner to interlocutor in SCMC followed the same patterns as in learner-system interaction, where alignment from learner to interlocutor typically occurred in the contexts of question-answer, leave-taking, or thanking adjacency pairs as well as reference shifts.

Example 128 shows an instance where the learner reuses “thank you very much” from their interlocutor’s previous message.

Example 128

Lexical alignment from interlocutor to learner in a thanking sequence in SCMC

35	Peer	<u>thank you very much</u> Dame Tahiri i'll book this for u!	<i>Interlocutor unprimed (First occurrence)</i>
36	Learner	<u>Thank you very much.</u>	<i>Learner to interlocutor</i>

Example 129 illustrates alignment from learner to interlocutor in the context of a question-answer pair. As can be seen, the learner reuses in message 7 the phrase “need a room” from their interlocutor’s production in message 6.

Example 129

Lexical alignment from interlocutor to learner in a question-answer pair in SCMC

6	Peer	Do you <u>need a room</u> ?	<i>Interlocutor unprimed (First occurrence)</i>
7	Learner	Good morning i <u>need a room</u> on the 12.09.2019	<i>Learner to interlocutor</i>

In Example 130, the learner in message 21 reuses the noun phrase “the newest technology” from the interlocutor’s production in message 15. As can be seen, the reference shifts from room Alpha to room Beta in message 17. So here, the alignment occurs in the context of a reference shift.

Example 130

Lexical alignment from interlocutor to learner involving a reference shift in SCMC

15	Peer	We got <u>the newest technology</u> and you also have a stage in there	<i>Interlocutor unprimed (First occurrence)</i>
16	Peer	The chairs are fix	
17	Learner	Good thank you and how is the room Beta?	
18 – 20		[...]	
21	Learner	Does this room also have <u>the newest technology</u> ? and how much does it cost?	<i>Learner to interlocutor</i>

Learner to Both

There were only two instances where the learner aligned to both in SCMC interaction. Whereas in learner-system interaction, alignment from learner to both occurred exclusively in the context of booking confirmations, the two instances in SCMC involved the phrase “in the room” or “in this room”, respectively. The two instances are presented in the examples below. In Example 131, the learner in message 14 reuses “in the room” that was previously used in the interlocutor’s production in message 5 as well as their own production in message 10.

Example 131

Lexical alignment from interlocutor to both in booking confirmations in SCMC

5	Peer	Yes of coure. <u>In the room</u> Alpha you can bring 270 people.	<i>Interlocutor unprimed (First occurrence)</i>
6 – 9		[...]	<i>Learner to interlocutor</i>
10	Learner	<u>in the room</u> Alpha is it possible to have a conversation in small groups?	
11 – 13		[...]	
14	Learner	and <u>in the room</u> Gamma is'nt possible to make a presation with videos and sound?	<i>Learner to both</i>

Similarly, in Example 132, the learner in message 17 reuses “in this room”, previously used by themselves (message 8) as well as their interlocutor (message 9).

Example 132

Lexical alignment from interlocutor to both in booking confirmations in SCMC

8	Learner	and we want to have an Presentation <u>in this room</u>	<i>Learner unprimed (First occurrence)</i>
9	Peer	I prefer the Gamma room max 200 person can stay <u>in this room</u>	<i>Interlocutor to learner</i>
10 – 16		[...]	
17	Learner	ok i have one last question, is it possible to have conversations in small groups <u>in this room</u> ?	<i>Learner to both</i>

Interlocutor to Learner

As shown above, in system interaction alignment from interlocutor to learner typically involved the confirmation of a booking or leave-taking and thanking formulas. In SCMC, these contexts were also found in a few instances of alignment from interlocutor to learner, as illustrated in the following examples.

Example 133 shows an instance where the interlocutor in message 9 reuses the phrase “a nice day” from the learner’s message 8 in a leave-taking sequence.

Example 133

Lexical alignment from interlocutor to learner in a leave-taking sequence in SCMC

8	Learner	Thanks for your service and have <u>a nice day</u> .	<i>Learner unprimed (First occurrence)</i>
9	Peer	Ok, we can arrange this. The One King Hotel wishes you <u>a nice day</u> .	<i>Interlocutor to learner</i>

More typically, however, alignment from interlocutor to learner in SCMC occurred in contexts where the learner's room requirement was repeated by the interlocutor. This pattern is illustrated in the following examples. In Example 134, the interlocutor in message 18 reuses the learner's phrase from message 16 "ta[l]k in small groups", relating to the room requirements.

Example 134

Lexical alignment from interlocutor to learner involving the repetition of a room requirement in SCMC

16	Peer	It doesn't have to be very big, it should give our inviters the oppurtunity to <u>talk in small groups</u> and it should show some good atmosphary	<i>Learner unprimed (First occurrence)</i>
17	Learner	The tables sound good	
18	Peer	So this room would be perfect for you and your familiy or friends. It has enough place to sit and <u>tak in small groups</u> .	<i>Interlocutor to learner</i>

In Example 135, the interlocutor reuses multiple of the learner's phrases describing the room requirements, i.e., "a relax situation", "can work in groups", and "device for recording and audio".

Example 135

Lexical alignment from interlocutor to learner involving the repetition of a room requirement in SCMC

28	Learner	i ask again, which one of these rooms are the best for 150 people, <u>a relax situation</u> , so that we <u>can work in groups</u> , nice	<i>Learner unprimed (First occurrence)</i>
----	---------	---	--

		impression and a <u>device for recording and audios</u> . the cost sholud'nt be over 1500 CHF	
29	Peer	the best room is alpha room, 270 people its a relax situation, they <u>can work in groups</u> and a nice impression and <u>device for recording and audio</u>	<i>Interlocutor to learner</i>

Similarly, in Example 136, the interlocutor reuses the phrase “to make a pres[ent]ation”.

Example 136

Lexical alignment from interlocutor to learner involving the repetition of a room requirement in SCMC

14	Learner	and in the room Gamma is'nt possible <u>to make a presation</u> with videos and sound?	<i>Learner unprimed (First occurrence)</i>
15	Peer	It have two display to <u>make a presentation</u> .	<i>Interlocutor to learner</i>

As can be seen in these examples, the interlocutor’s reuse may repeat a learner error (Example 135), contain a new error (Example 134), or correct an error (Example 136). This is similar to learner-system interaction where alignment from interlocutor to learner sometimes involved error correction in the context of recasts or error repetition in the context of booking dates.

Learner Unprimed

As presented above, in learner-system interaction, learners’ unprimed reuse of n-grams typically involved reference shifts, rephrasing, or thanking routines. In SCMC, learners’ unprimed reuse of lexical material was found much less frequently (as shown in the overview in Table 44), however, in similar contexts. Example 137 shows an example of unprimed learner reuse of “does the room” involving a shift in reference from relating to Alpha in message 14 to room Gamma in message 25.

Example 137*Unprimed learner reuse involving a change of reference in SCMC*

14	Learner	and what equipment <u>does the room</u> have?	<i>Learner unprimed (First occurrence)</i>
15 – 24		[...]	
25	Learner	What <u>does the room</u> Gamma offer?	<i>Learner unprimed</i>

Example 138 illustrates an instance where the learner rephrases a previous message. Here, the learner rephrases the initial booking date statement in message 21 “for 12 of october 2019” as “Is it free at the 12 of october”, thus reusing the date “12 of october”.

Example 138*Unprimed learner reuse involving rephrasing in SCMC*

21	Learner	for <u>12 of october</u> 2019	<i>Learner unprimed (First occurrence)</i>
22 – 31		[...]	
32	Learner	Is it free at the <u>12 of october</u> ?	<i>Learner unprimed</i>

Similarly, in Example 139, the learner in message 14 rephrases and expands the utterance from message 3, reusing the phrase “be big enough for”. Note that “The room must” is not coded as *unprimed* reuse because the item was used by the learner in message 8 (not shown here).

As can be seen in these examples, the learner’s rephrasing involving unprimed learner reuse in SCMC was not prompted by the interlocutor’s indication of non-understanding. This contrasts with learner-system interaction where it was typically preceded by the interlocutor’s indication of non-understanding.

Example 139*Unprimed learner reuse involving rephrasing in SCMC*

3	Learner	The room must <u>be big enough for</u> 150 people and the price limit is about 1'500 francs. Is that possible?	<i>Learner unprimed (First occurrence)</i>
---	---------	--	--

4 – 13		[...]	
32	Learner	So know import for us is that: The room must have a projector, have to <u>be big enough for</u> about 150 people, conversations in small groups must be possible, a nice atmosphere, not too much, price limit 1500 francs	<i>Learner unprimed</i>

Example 140 illustrates unprimed learner use involving thanking.

Example 140

Unprimed learner reuse as part of a thanking sequence in SCMC

17	Learner	<u>Good thank you</u> and how is the room Beta?	<i>Learner unprimed (First occurrence)</i>
18 – 39		[...]	
40	Learner	<u>Good thank you</u>	<i>Learner unprimed</i>

Interlocutor Unprimed

As will be seen in the overview below, unprimed reuse by the interlocutor was very rare in SCMC. As in learner-system interaction, the few instances of unprimed interlocutor reuse in SCMC occurred in the context of booking dates and task-related vocabulary. The other contexts found in system interaction such as room names and leave-taking/thanking did not occur in the SCMC data.

In Example 141, the interlocutor in message 39 reuses the booking date “for the 12 of october” from message 21.

Example 141

Unprimed interlocutor reuse as part of a booking confirmation in SCMC

21	Learner	<u>for the 12 of october</u> 2019	<i>Learner unprimed (First occurrence)</i>
22 – 38		[...]	
39	Learner	Ok we will make the room ready for you <u>for the 12 of october</u>	<i>Interlocutor unprimed</i>

Example 142 shows an instance where the interlocutor reuses the task-related verb phrase “do you have” to form a question.

Example 142

Unprimed interlocutor involving task-related vocabulary in SCMC

5	Learner	<u>Do you have</u> one?	<i>Learner unprimed (First occurrence)</i>
6 – 36		[...]	
37	Peer	<u>do you have</u> any other question?	<i>Interlocutor unprimed</i>

As in learner-system interaction, unprimed interlocutor reuse may occur within closer proximity to the first occurrence (e.g., within 18 messages as in Example 141) as well as with larger distance (e.g., within 32 messages as in Example 142).

Learners' Perception on Lexical Alignment

During the interviews, alignment was only mentioned once by a learner in each group. The excerpt from the system group is presented below. In this instance, the learner indicated that they actively looked up and copied an unknown adequate term from the interlocutor's production:

weil er "Beamer" nicht verstand. Und dann ging ich schauen, ob er schon einmal geschrieben hatte, [...] was ein Beam-- [...] wie heisst er, äh-- Projektor oder-- wie das halt auf Deu-- äh, auf Englisch heisst.

The learner from the SCMC group explained that they reused a lexical item from the interlocutor's production to enhance mutual understanding:

weil ich übernahm das genauso, wie sie es vorschrieb. Da wusste sie sicher auch gleich sofort, [...] was ich meine. Weil ich hatte ja vorher die Frage gestellt wegen dem "equipment", da antwortete sie damit.

4.2.4 Structural Alignment

As explained previously (see section Coding of Structural Alignment in the chapter Methods), structural alignment refers to the reuse of a syntactical structure

that occurred previously in the interaction, either in the learner's own or the interlocutor's production. As also shown above (see section Coding of Structural Alignment in the chapter Methods), for analysis, the present study considered frequent syntactical structures that occurred at least five times within a given interaction and that were found in both the interlocutor's and learner's production. The frequent syntactical structures falling under this definition that were found in the data are presented again along with examples in Table 45. The table also lists the number of total occurrences per structure for each interaction type. Total occurrences include all instances of a given structure, regardless of whether they were subsequently coded as *structural alignment*, i.e., the reuse of a structure within the threshold of five messages (system) or 10 messages (SCMC), or as *unprimed*, i.e., the reuse of a structure beyond the threshold (see section Coding of Structural Alignment in the chapter Methods). Occurrences per message correct the measure for text length and allow for better comparability between the two interaction types.

As this overview shows, all structures except *have* + noun phrase occurred more frequently in learner-system interaction. This pattern is particularly pronounced for question with subject-auxiliary inversion, which occurred more than five times more frequently in learner-system interaction and existential *there*, which did not occur in SCMC at all.

Table 45

Frequencies of Syntactical Structures in Learner-System Interaction and SCMC

Structure	System		SCMC	
	Total	Per message (<i>n</i> = 1139)	Total	Per message (<i>n</i> = 611)
Copula <i>be</i>	484	0.425	236	0.386
<i>Have</i> + noun phrase	183	0.161	116	0.190
<i>For</i> + noun phrase	384	0.337	111	0.182
Modal verb + infinitive	489	0.429	113	0.185
<i>To</i> -infinitive	139	0.122	44	0.072

Structure	System		SCMC	
	Total	Per message (<i>n</i> = 1139)	Total	Per message (<i>n</i> = 611)
Existential <i>there</i>	25	0.022	-	-
Question with subject-auxiliary inversion	273	0.240	27	0.044
<i>Wh</i> -interrogative	103	0.090	35	0.057

Structural Alignment in Learner-System Interaction

Table 46 presents an overview of the total, per interaction, and per message frequencies of the different alignment types found in learner-system interaction across all structures. As the overview shows, total structural alignment was 1345 or 64.048 per interaction. Most learner reuse of structures was within-threshold alignment (502 or 23.905 per interaction), whereas unprimed reuse was relatively infrequent (79 or 3.762 per interaction). The most frequent type of learner alignment was from learner to interlocutor (215 or 10.238 per interaction), followed by alignment from learner to both (182 or 8.677 per interaction), and from learner to learner (self) with about half as many occurrences (105 or 5.000 per interaction). Interlocutor reuse across all structures (764 or 36.381 per interaction) was higher than learner reuse. Most interlocutor reuse occurred within the threshold at 548 (26.095 per interaction), while the remaining occurrences of interlocutor reuse were unprimed (216 or 10.286 per interaction).

Table 46

Frequency of Subcodes for Structural Alignment Across all Structures in Learner-System Interaction

Subcode	Total count	Per interaction (<i>n</i> = 21)	Per message (<i>n</i> = 1139)
Used by learner	581	27.667	0.510
Learner to learner	105	5.000	0.092
Learner to interlocutor	215	10.238	0.189
Learner to both	182	8.667	0.160
Total learner alignment	502	23.905	0.441

Subcode	Total count	Per interaction (<i>n</i> = 21)	Per message (<i>n</i> = 1139)
Learner unprimed ^a	79	3.762	0.069
Used by interlocutor	764	36.381	0.671
Interlocutor to learner	548	26.095	0.481
Interlocutor unprimed ^a	216	10.286	0.190
Total	1345	64.048	1.181

^a *Unprimed* counts exclude the first occurrences of structures in an interaction, although first occurrences were included in the *unprimed* code during coding for coding consistency reasons. To obtain the *unprimed* counts excluding first occurrences, all first occurrences were deducted by hand from the *unprimed* count in MAXQDA.

Structural Alignment in SCMC

To compare the structural alignment between the two interaction types, the same structures were identified and coded in SCMC. As mentioned above, the threshold in SCMC was adjusted to 10 messages (from five messages in learner-system interaction) to account for the different chat interfaces. The findings revealed that structural alignment in SCMC involved the same alignment directions as learner-system interaction.

Table 47 summarizes the occurrences of each alignment type. The table shows the overall count as well as the occurrences per interaction and per message to correct for different numbers of interactions and messages between the two interaction types. As can be seen in the table, total structural reuse in SCMC uncorrected for number of interactions and messages was about 2.5 times lower than in learner-system interaction (1345 system; 545 SCMC). Per message, total reuse in SCMC was still 1.3 times lower than in learner-system interaction (1.181 system; 0.892 SCMC).

Focusing on learner reuse, total learner reuse of structures (i.e., within-threshold learner alignment and unprimed learner reuse) uncorrected for number of messages was 2.1 times lower in SCMC than in learner-system interaction (581 system; 280 SCMC). However, learner reuse per message in SCMC was almost equal with learner-system interaction (0.510 system; 0.458 SCMC). As in learner-system interaction, learner reuse of structures in SCMC predominantly occurred within-threshold. However, the proportion of unprimed learner reuse in SCMC was

considerably lower than in learner-system interaction: In SCMC, there was one unprimed learner reuse for every 30 within-threshold learner alignments. In learner-system interaction, there was one unprimed learner reuse for every about six within-threshold learner alignments.

The majority of the within-threshold learner alignment in SCMC involved alignment from learner to both (159 or 8.833 per interaction or 0.260 per message), followed by alignment from learner to learner (self) (69 or 3.833 per interaction or 0.113 per message), and from learner to interlocutor (43 or 2.389 per interaction or 0.070 per message). This distribution contrasts with learner-system interaction, where alignment from learner to interlocutor and from learner to both were both about twice as frequent as alignment from learner to learner (self). Comparing the occurrence of each individual direction of learner alignment between the two interaction types, SCMC had lower alignment from learner to interlocutor than learner-system interaction (system: 0.189 per message; SCMC: 0.070 per message), but SCMC had higher alignment from learner to learner (self) (system: 0.092 per message; SCMC: 0.113 per message) and higher alignment from learner to both (system: 0.160 per message; SCMC: 0.260 per message).

With respect to interlocutor reuse, total interlocutor reuse of structures (i.e., within-threshold interlocutor alignment and unprimed interlocutor reuse) was lower in SCMC than in learner-system interaction both uncorrected for number of messages (system: 764; SCMC: 265) as well as per message (system: 0.671; SCMC: 0.434). As in learner-system interaction, within-threshold interlocutor alignment in SCMC was higher than unprimed interlocutor reuse (255 within threshold; 10 unprimed). As with learner unprimed reuse vs. within-threshold alignment, the proportion of interlocutor unprimed reuse to within-threshold interlocutor alignment was lower in SCMC compared to that in learner-system interaction: In SCMC, there was one unprimed interlocutor reuse for every 25.5 within-threshold interlocutor alignments, whereas in learner-system interaction, there was one unprimed reuse for every 2.5 interlocutor alignments.

Table 47

Frequencies of Subcodes for Structural Alignment Across all Structures in Learner-System Interaction and SCMC

Subcode	System			SCMC		
	Total count	Per interaction (<i>n</i> = 21)	Per message (<i>n</i> = 1139)	Total count	Per interaction (<i>n</i> = 18)	Per message (<i>n</i> = 611)
Used by learner	581	27.667	0.510	280	15.556	0.458
Learner to learner	105	5.000	0.092	69	3.833	0.113
Learner to interlocutor	215	10.238	0.189	43	2.389	0.070
Learner to both	182	8.667	0.160	159	8.833	0.260
Total learner alignment	502	23.905	0.441	271	15.056	0.444
Learner unprimed ^a	79	3.762	0.069	9	0.500	0.015
Used by interlocutor	764	36.381	0.671	265	14.722	0.434
Interlocutor to learner	548	26.095	0.481	255	14.167	0.417
Interlocutor unprimed ^a	216	10.286	0.190	10	0.556	0.016
Total	1345	64.048	1.181	545	30.278	0.892

^a *Unprimed* counts exclude the first occurrences of structures in an interaction, although first occurrences were included in the *unprimed* code during coding for coding consistency reasons. To obtain the *unprimed* counts excluding first occurrences, all first occurrences were deducted by hand from the *unprimed* count in MAXQDA.

Learners' perception on structural alignment

No additional data on structural alignment were gained from the questionnaires and interviews, as learners did not comment on it.

4.2.5 Summary of Findings on Research Question 1

This subsection presents an initial summary of the findings in response to research question 1 on the occurrence and characteristics of interactional features in learner-system interaction compared to learner-peer SCMC. The findings will be summarized, discussed, and interpreted in detail in the next chapter (see chapter Discussion).

Overall, the analysis of the learner-system interaction scripts yielded the following interactional features: negotiation of meaning, recasts, and lexical and structural alignment. Each feature was further analyzed by assigning subcodes to their specific characteristics. Some subcodes could be taken from the literature on human-human interaction while others had to be developed new from the data, which indicates that interactional features occurring in the learner-system interaction data might have different characteristics than those in learner-peer SCMC. When applying the same subcodes to the SCMC data, overlap and differences in the presence and realization of the subcodes revealed to what degree the features' characteristics matched between the two interaction types.

With regard to the interactional feature *negotiation of meaning*, it was found that the feature occurred much more frequently in learner-system interaction compared to SCMC. The subcodes on the individual negotiation components (trigger, indicator, response, reaction to response) and the parallels and differences in their realization and counts between the two interaction types further revealed their interaction-specific characteristics: Focusing on the trigger and indicator components, non-understanding in learner-system interaction was almost always indicated by the interlocutor (i.e., dialog system), whereas in SCMC, indicators were more equally used by learner and interlocutor (i.e., peer), albeit they were very infrequent overall, as mentioned above. As a parallel between the two interaction types, the same indicator types (clarification request, confirmation check, inappropriate response) could be

observed in learner-system interaction and SCMC, although inappropriate responses were relatively considerably more frequent in learner-system interaction than in SCMC, and the linguistic realization of clarification requests was strikingly more uniform in learner-system interaction than in SCMC. Furthermore, indicators in both interaction types generally failed to point to a specific element in the trigger message that caused non-understanding and instead referred to the entire message content. Similarly, in both interaction types, both erroneous and target-like triggers led to indications of non-understanding to about equal measures. However, a characteristic specific to learner-system interaction was that indicators generally related to content-related issues, whereas SCMC indicators tended to point to task-related issues (i.e., clarifying task roles).

With regard to the response component, the learner-system subcodes were found to match the responses occurring in the SCMC data, which indicates a degree of overlap between the two interaction types. However, not all learner-system response subcodes were present in SCMC, presumably partly due to the overall low occurrence of responses in SCMC. As in learner-system interaction, SCMC may involve 'non-responses' such as *meta-comment* and *accept topic*. However, the 'non-response' type *meta-comment* was realized differently in SCMC in that it managed to resolve the non-understanding, whereas meta-comments in learner-system interaction served to exit the negotiation sequence. Most response types acknowledging the indicated non-understanding which could be observed in learner-system interaction were absent from the SCMC data, with the exception of the type *elaboration* (i.e., *reduction*, *modification*, *repetition* did not occur).

With regard to the final component *reaction to response*, there was a noticeable difference between the two interaction types: In SCMC, reaction to response always involved an explicit statement that the non-understanding had been resolved. In learner-system interaction, on the other hand, this reaction type never occurred. Instead, learner-system interaction involved implicit indications that the non-understanding had been resolved or, respectively, involved implicit or explicit indications that the non-understanding had not yet been resolved.

A final prominent difference between the two interaction types was the absence of multi-layered negotiation sequences in SCMC. In learner-system interaction, on the

other hand, they occurred frequently and involved up to four layers of overlapping negotiation sequences.

Summarizing the findings on the learners' perception on negotiation of meaning gained from the supplementary data, the questionnaire findings indicated that most learners from both groups perceived their understanding of the interlocutor as high. However, when asked about how well the interlocutor understood them, about half of the learners from the system group reported low understanding, while the majority of learners from SCMC indicated no understanding issues. The interview findings additionally revealed that most learners interacting with the dialog system felt a range of positive and negative emotions in response to the lack of understanding, ranging from curiosity and amusement at the positive end to frustration/annoyance, losing patience, and confusion at the negative end.

With respect to the reported strategies that learners used to deal with non-understanding on the part of the interlocutor, learners from groups reported in the questionnaire to primarily having used non-interactive strategies such as *deriving meaning from context* or *checking in chat*. To negotiate understanding when learners themselves were not understood, learners from the system group reported considerably more and more diverse strategies than those from the SCMC group. The findings from the interviews additionally indicated that most interviewed learners interacting with the dialog system considered both themselves as well as the dialog system responsible for these instances of non-understanding. Learners furthermore specified how they thought their respective behaviors contributed to the non-understanding. While some of these conceptions accurately describe the processes leading to a lack of understanding by the dialog system (e.g., the dialog system is often unable to process long messages) others were only partly accurate (e.g., the dialog system cannot retain information; the dialog system relies on keyword extraction). The interview findings further revealed that the learners often chose a strategy to deal with the non-understanding (e.g., repeat their message; use a single keyword) that corresponded to their conception of what created the non-understanding.

As a general observation on the interactional feature *recast*, not all subcodes pertaining to recasts from learner-system interaction could also be applied to SCMC.

Thus, absent from the SCMC data were the linguistic focus subtype *lexis, repeated incidence [of recast within interaction], complete recast, immediate uptake/repair, delayed uptake/needs repair, and re-uptake*. This lack of overlap serves as an initial indication of some essential differences in the characteristics of the feature *recast* between the two interaction types. At the same time, no further subcodes had to be developed to code the SCMC recasts, which conversely points to a degree of similarity between the two interaction types.

Considering the findings on the counts and realizations of each recast subcode, further similarities as well as differences between the two interaction types could be observed: With regard to the linguistic focus of the recasts, in both interaction types, the *grammar* subcode was the most prevalent. As a further qualitative finding, recasts of the subtype *grammar* in the two interaction types generally also focused on the same error types (e.g., omission of indefinite article, incorrect inflection for person of nouns, misselection of words). As noted above, the subcode *lexis*, found with moderate frequency in learner-system interaction, was absent from the SCMC data, pointing to a marked difference between the two interaction types. The subcode *spelling* was relatively more frequent in SCMC compared to learner-system interaction. However, both interaction types addressed the same types of spelling errors (e.g., omission or misselection of letters, typing errors). The subcode *multi-focus* was relatively more frequent in SCMC; however, the occurrence was low overall (two occurrences in both learner-system interaction and SCMC).

With respect to the recast intensity, about half of those observed in learner-system interaction were of the subtype *repeated incidence [of recast within interaction]*, whereas all SCMC recasts were coded as *singular incidence*. However, repetition of the recast form within the same recast move occurred in both interaction types, with a slightly higher relative frequency in SCMC.

As noted above, none of the SCMC recasts were coded as *complete*, whereas almost half of system recasts were complete. This indicates a notable difference in the degree to which recasts in the two interaction types left errors unaddressed in partial recasts.

All subcodes pertaining to the length of the recast found in learner-system interaction were observed in SCMC as well; however, there was a difference in the

relative prevalence of the different subtypes: In SCMC recasts were mostly embedded in multiple messages and system recasts mostly in single sentences.

With regard to the subcode *uptake*, SCMC contained one immediate uptake, one delayed uptake, and no re-uptake at all. This contrasted with the learner-system data, where half of the recasts were followed by immediate or delayed uptake of which again about half involved re-uptake. Due to the overall low occurrence of SCMC uptake, the degree of repair could not be measured conclusively for this interaction type. In system interaction, 60.0% of immediate and delayed uptake of the recast form contained repair whereas 40.0% were still in need of repair. As a further observation about uptake in system recasts, immediate uptake appeared to lead to repair more successfully than delayed uptake which was mostly still in need of repair.

With respect to lexical alignment, the absolute number of occurrences uncorrected for unequal number of interactions and messages in the two interaction types revealed that learner-system interaction produced almost three times more lexical reuse overall (including within-threshold alignment and unprimed reuse) compared to SCMC, and from 1.5 to ten times more with respect to the individual alignment directions. However, corrected for number of messages, the differences between the two interaction types were less marked or even reversed: Per message, learner-system interaction still had higher total lexical reuse (i.e., within-threshold alignment and unprimed reuse) and slightly higher total within-threshold learner alignment, as well as alignment from learner to learner (self), and from learner to both. However, SCMC had higher per message alignment from learner to interlocutor and from interlocutor to learner. In addition, both unprimed learner reuse and interlocutor reuse were considerably higher in learner-system interaction than in SCMC.

Focusing on the distribution of the different alignment directions within each interaction type showed that both interaction types led to more learner than interlocutor lexical alignment. Focusing on learner alignment, alignment from learner to learner (self) was the most frequent within-threshold alignment direction in both interaction types, followed by alignment from learner to interlocutor, and the least frequent alignment from learner to both. However, as a noticeable difference between

the two interaction types, in learner-system interaction, there was a clear prevalence of alignment from learner to learner (self) compared to the other two types of learner alignment, whereas in SCMC the prevalence of alignment from learner to learner (self) was much less pronounced. As another difference between the two interaction types, in learner-system interaction, unprimed learner reuse of lexical material was as frequent as within-threshold learner alignment, but in SCMC it was relatively infrequent compared to within-threshold learner alignment.

The two directions of interlocutor reuse of lexical material (i.e., within-threshold alignment from interlocutor to learner and unprimed interlocutor reuse) distributed similarly in the two interaction types, with within-threshold interlocutor-to-learner alignment being considerably more frequent than unprimed interlocutor reuse.

The qualitative findings revealed parallels and differences with respect to the linguistic contexts in which the different directions of lexical alignment occurred in system interaction and SCMC. In learner-system interaction, alignment from learner to learner (self) may involve the learner quasi-repeating a previous request while shifting its reference, or it may occur as part of a negotiation sequence. In SCMC, alignment from learner to learner (self) was also found in the context of reference shifts, but not in the context of negotiation of meaning. Instead, alignment from learner to learner (self) in SCMC may involve the learner restating, specifying, or correcting their previous statements without being prompted to do so by the interlocutor's indication of non-understanding. The next direction of learner alignment, from learner to interlocutor, in both interaction types typically involved adjacency pairs (e.g., question-answer, thanking, and leave-taking) or, rarely, quasi-repetitions with reference shifts. Alignment from learner to both in learner-system interaction exclusively involved booking confirmations. In SCMC, this direction only occurred twice and involved "in the room" and "in this room", respectively. Finally, unprimed learner reuse of lexical material was found in similar contexts in the two interaction types, albeit it was much less frequent overall in SCMC than in learner-system interaction, involving reference shifts, rephrasing, or thanking routines. However, as in alignment from learner to learner (self) described above, unprimed

learner reuse of lexical material was not prompted by the interlocutor's indication of non-understanding as it was in learner-system interaction.

Focusing on the different directions of interlocutor alignment combined with the contexts in which they occurred, alignment from interlocutor to learner in both interaction types may involve the confirmation booking requests or leave-taking and thanking-expressions. SCMC alignment from interlocutor to learner additionally involved the context where the interlocutor repeated a room requirement that the learner had previously stated. Unprimed interlocutor reuse of lexical material was very rare in SCMC. The few observed instances occurred in the contexts of booking dates and task-related vocabulary (e.g., "do you have"), which were also found in learner-system unprimed interlocutor reuse of lexical material. The other contexts observed in learner-system interlocutor unprimed reuse, however, i.e., room names, leave-taking/thanking, were not present in the SCMC data.

Structural alignment was analyzed for eight frequent grammatical structures in learner-system interaction and SCMC. The total occurrences of the analyzed structures (including unprimed reuse, all directions of within-threshold alignment, as well as interlocutor-to-interlocutor alignment otherwise excluded from the analysis) were generally higher in learner-system interaction than in SCMC, both when uncorrected and corrected for number of messages. One exception was *have* + noun phrase which had a higher total occurrence in SCMC. Also notable was the absence of the reuse of existential *there* in the entire SCMC data.

Comparing the occurrences in the two interaction types, total structural reuse (i.e., within-threshold alignment, unprimed reuse) uncorrected for number of messages across all structures was about 2.5 times higher in learner-system interaction than in SCMC. Corrected for number of messages, the difference was less pronounced, with learner-system interaction producing 1.3 times more total alignment per message. Similarly, with respect to learner reuse (i.e., within-threshold learner alignment and unprimed learner reuse), learner-system interaction produced 2.1 times more overall reuse and 1.1 times more reuse per message. Total interlocutor reuse (i.e., within-threshold interlocutor alignment and unprimed interlocutor reuse) across all

directions uncorrected for number of messages was 2.9 times higher in learner-system interaction than in SCMC and 1.5 times higher per message.

With respect to how learner reuse distributed across the various directions, there were both similarities and differences between the two interaction types: In both learner-system interaction and SCMC, learners reused structures predominantly within threshold rather than unprimed. However, the ratio of unprimed learner reuse to within-threshold learner alignment was considerably lower in learner-system interaction compared to SCMC (1 : 6 in system; 1 : 30 in SCMC). As another difference, in learner-system interaction, alignment from learner to interlocutor and from learner to both were the dominant directions, with alignment from learner to learner (self) occurring about half as often compared to the other two directions. In SCMC, alignment from learner to learner (self) and from learner to both were the most frequent directions, while alignment from learner to interlocutor was less frequent.

With respect to interlocutor reuse of structures, there was proportionally more within-threshold alignment than unprimed interlocutor reuse in both interaction types. However, the ratio of within-threshold to unprimed interlocutor reuse of structures was again considerably lower in learner-system interaction than in SCMC (1 : 2.5 in system; 1 : 25.5 in SCMC).

Since alignment was only mentioned once in each group during interviews, the findings on the learners' perception of this interactional feature are limited. The learner from the system group indicated that they actively used a term that was lacking in their own vocabulary from the dialog system's production. The learner from the SCMC group reused a lexical item from the interlocutor's production to ensure to be understood by them.

4.3 Research Question 2: Learners' Task Performance

This section reports the findings on research question 2, i.e., the comparison of the linguistic complexity and accuracy as well as communicative adequacy of the learners' task performance in the two interaction types. As described in the chapter Methods (see section

Measures for Complexity, Accuracy, and Communicative Adequacy), a range of measures were selected to evaluate the learners' task performance with respect to the dimensions above. As also described in the chapter Methods, the measures were established for each interaction type and subsequently tested for statistical differences between system interaction and SCMC. The results from these statistical analyses will be presented in the following section and supplemented with descriptive findings whenever conducive. As mentioned above, the findings will be interpreted and discussed in the chapter Discussion.

4.3.1 Complexity

The descriptive statistics along with the results from the statistical tests on the linguistic complexity measures in learner-system interaction and SCMC are presented in Table 48.

T-tests were conducted on all variables that met the conditions for this statistical test (such as interval scale, normal distribution, and variance homogeneity). Preliminary Shapiro-Wilk tests at level $\alpha = .05$ detected no significant deviation from normality for any of the variables. Due to significant lack of variance homogeneity in the variable *words per AS-unit*, the Welch's test was chosen to assess group differences in this variable.

As seen in Table 48, the results indicate significantly lower lexical diversity as measured by MTLT in learner-system interaction ($M = 45.33$, $SD = 13.41$) than in SCMC ($M = 56.06$, $SD = 16.10$) with a medium to large effect size ($t(37) = -2.27$, $p = .029$, $d = -0.72$). Learners also produced significantly fewer words per AS-unit in learner-system interaction ($M = 5.48$, $SD = 1.29$) compared to SCMC ($M = 8.42$, $SD = 4.15$) representing a large effect size ($t(37) = -2.88$, $p = .009$, $d = -0.95$). However, the interaction type had no effect on the syntactic similarity of learners' production as measured by the SYNSTRUTt index (system: $M = 0.10$, $SD = 0.03$; SCMC: $M = 0.10$, $SD = 0.03$; $t(37) = 0.24$, $p = .812$, $d = 0.08$).

Table 48

Descriptive Statistics and Results of Independent Sample T-Test (Two-tailed) for Differences in Complexity Measures Between Learner-System Interaction and SCMC

Measure	<i>M (SD)</i>		<i>t</i>	<i>p</i>	Cohen's <i>d</i>
	System (<i>n</i> = 21)	SCMC (<i>n</i> = 18)			
MTLD ^a	45.33 (13.41)	56.06 (16.10)	-2.27	.029*	-0.72
Words per AS-unit ^{b, c}	5.48 (1.29)	8.42 (4.15)	-2.88**	.009**	-0.95
SYNSTRUT ^d	0.10 (0.03)	0.10 (0.03)	0.24	.812	0.08

^a MTLD = Measure of Textual Lexical Diversity, for calculation see section Lexical Diversity Measured by MTLD. ^b AS-unit = Analysis of Speech Unit, for definition and coding see section Coding of AS-Units.

^c Due to significant ($p < .05$) lack of variance homogeneity, results from the Welch's test are reported for this variable. ^d SYNSTRUT = Index of sentence syntax similarity between all sentences within a text, for definition and calculation see section Syntactic Similarity Measured by SYNSTRUT.

* $p < .05$. ** $p < .01$.

Learners' Perception on Complexity

Learners commented on the complexity of their production during interviews. An overview of the derived codes pertaining to complexity can be found in the appendix (Appendix L). The comments included statements about the length of messages and clauses or the overall simplicity or complexity, respectively, of the formulation. The statements commented on complexity at the level of various linguistic units: Statements relating to the message-level primarily focused on the number of requests that learners added per message, i.e., the question whether each message should contain a single request to reduce the content load of each message or could include multiple requests. At the sentence-level, learners focused on sentence length or the general complexity of a sentence. Complexity at the word-level, i.e., lexical diversity, was only mentioned in the SCMC group. In the remaining statements, the learner did not specify the aspect on which their statement focuses.

Learners indicated a range of reasons to account for the complexity of their production which could be classified into a number of subcodes (for an overview with

examples, see Appendix M). In the system group, learners increased the complexity to model their production after an email or to increase the information content of a message. In the SCMC group, learners indicated that they increased the complexity with the goal of applying Business English to enhance learning. In one instance, the learner did not specify the reason for enhancing the complexity of their production.

As reasons for reducing the complexity of their production, learners in both groups indicated that they lacked the competence or chose to reduce complexity to enhance understanding. In the system group, one learner additionally stated that they lost patience to produce complex language. In the SCMC group, a learner indicated that they reduce the complexity due to the pressure to reply quickly. Relating to complexity at the message-level, one participant furthermore mentioned that sending additional messages before having received a response to the first message was considered bad chatting etiquette.

4.3.2 Accuracy

The descriptive statistics of the linguistic accuracy measures in learner-system interaction and SCMC are presented in Table 49 along with the results of the statistical tests assessing differences between the two interaction types.

Preliminary Shapiro-Wilk tests at level $\alpha = .05$ detected no significant deviation from normality for any of the variables. Since the variable *request appropriateness* is ordinally scaled, the Mann-Whitney U test was chosen to assess group differences for this variable.

While statistical analysis revealed small effects for three of the four accuracy measures, none of them reached statistical significance. Thus, learners produced fewer errors per AS-unit in learner-system interaction ($M = 0.66$, $SD = 0.31$) than in SCMC ($M = 0.72$, $SD = 0.27$), corresponding to a small, non-significant effect size ($t(37) = -0.71$, $p = .482$, $d = -0.21$). Accordingly, learners also produced more error-free AS-units per total AS-units in learner-system interaction ($M = 0.61$, $SD = 0.13$, SCMC) than in SCMC ($M = 0.56$, $SD = 0.31$) with a small non-significant effect ($t(37) = 1.14$, $p = .260$, $d = 0.37$).

However, with respect to the appropriateness of requests, learners interacting with the dialog system produced more inappropriate requests per total requests ($M = 0.26$, $SD = 0.20$) than SCMC ($M = 0.19$, $SD = 0.19$), corresponding to a small effect that did not reach statistical significance ($U = 155.00$, $p = .334$, $d = 0.34$). Learners' mean request appropriateness scores were also slightly lower in learner-system interaction ($M = 1.96$, $SD = 0.40$) than in SCMC ($M = 2.01$, $SD = 0.29$), although this effect was negligible in size and non-significant ($U = 160.50$, $p = .422$, $d = -0.13$).

Table 49

Descriptive Statistics and Results of Independent Sample T-Test (Two-tailed) and Mann-Whitney U Test for Differences in Accuracy Measures Between Learner-System Interaction and SCMC

Measure	<i>M (SD)</i>		<i>t</i>	Mann-Whitney <i>U</i>	<i>p</i>	Cohen's <i>d</i>
	System (<i>n</i> = 21)	SCMC (<i>n</i> = 18)				
Errors per AS-unit ^a	0.66 (0.31)	0.72 (0.27)	-0.71	-	.482	-0.21
Error-free AS-units per AS-unit ^a	0.61 (0.13)	0.56 (0.13)	1.14	-	.260	0.37
Inappropriate requests per request ^b	0.26 (0.20)	0.19 (0.19)	-	155.00	.334	0.34
Mean request appropriateness score ^c	1.96 (0.40)	2.01 (0.29)	-	160.50	.422	-0.13

^a AS-unit = Analysis of Speech Unit, for definition and coding see section Coding of AS-Units. ^b Due to non-normally distributed data, the Mann-Whitney *U* test was conducted for this variable. ^c Due to ordinal-scaled data, the Mann Whitney *U* test was conducted for this variable.

The following descriptive overviews on error types and request appropriateness serve to further describe and qualitatively compare the linguistic accuracy found in the two interaction types. As Table 50 on error types shows, the proportions of lexical errors in learner-system interaction and SCMC differed only marginally (system: 28.4%; SCMC: 29.5%). However, the proportion of grammatical errors was higher in learner-system interaction compared to SCMC (system: 41.9%; SCMC: 31.7%), while the proportion of spelling errors was lower (system: 29.7%; SCMC: 38.7%).

Table 50

Frequency and Distribution of Subcodes for Types of Errors in Learner-System Interaction and SCMC

Subcode	System		SCMC	
	Count	%	Count	%
Grammar	168	41.9	86	31.7
Spelling	119	29.7	105	38.7
Lexis	114	28.4	80	29.5
Total	401	100.0	271	100.0

As seen in Table 51 on request appropriateness, the proportions of fully appropriate requests were slightly higher in learner-system interaction and SCMC (system: 20.3%; SCMC: 18.6%). The proportion of inappropriate requests was also higher in learner-system interaction than in SCMC (system: 26.6%; SCMC: 20.5%). However, learner-system interaction led to proportionally fewer partly appropriate requests (system: 53.1%; SCMC: 60.9%).

Table 51

Frequency and Distribution of Subcodes/Levels for Appropriateness Ratings of Requests in Learner-System Interaction and SCMC

Subcode/Level	System		SCMC	
	Count	%	Count	%
Inappropriate	72	26.6	32	20.5
Partly appropriate	144	53.1	95	60.9
Fully appropriate	55	20.3	29	18.6
Total	271	100.0	156	100.0

In addition, Table 52 displays the different request strategies used in each interaction type along with their absolute and relative frequencies, ordered by highest frequency in system interaction. As seen in the table, in both interaction types, learners predominantly used direct questions (system: 43.5%; SCMC: 41.7%) and strong hints (system: 28.8%; SCMC: 30.8%) to formulate requests, with learners in learner-system interaction relying slightly more on direct questions than those interacting through SCMC, while learners in SCMC used strong hints slightly more often than learners interacting with the dialog system.

Elliptical question (system: 8.1%; SCMC: 10.3%), *want* statement (system: 8.5%; SCMC: 6.4%), and reference to preparatory condition (system: 5.5%; SCMC: 3.8%) were used with medium frequency in both interaction types. While *want* statements and references to preparatory condition were slightly more frequent in system interaction, the proportion of elliptical questions was slightly higher in SCMC.

Imperative (system: 1.9%, SCMC: 0.6%) and indirect question (system: 1.9%; SCMC: 0.6%) were the least frequent types in both interaction types, with learner-system interaction showing a proportionally higher use of both imperatives and indirect questions compared to SCMC. While assertion belonged to the least frequently used strategies in system interaction, it was used with medium frequency compared to the other strategies in SCMC (system: 1.9%; SCMC: 5.8%).

Table 52

Frequency and Distribution of Subcodes for Type of Request Strategy in Learner-System Interaction and SCMC

Subcode	System		SCMC	
	Count	%	Count	%
Direct question	118	43.5	65	41.7
Strong hint	78	28.8	48	30.8
Elliptical question	22	8.1	16	10.3
<i>Want</i> statement	23	8.5	10	6.4
Reference to preparatory condition	15	5.5	6	3.8
Imperative	5	1.8	1	0.6
Indirect question	5	1.8	1	0.6
Assertion	5	1.8	9	5.8
Total	271	100	156	100

Table 53 presents the frequencies of the different types of modifiers used in requests, ordered by highest frequency in learner-system interaction. As seen in the table, modal preterite and agreeer were the dominant modifier types in both interaction types, accounting for almost two thirds of all modifiers used in both learner-system interaction and SCMC, followed by grounder accounting for 10.4% in the system group and 17.1% in the SCMC group.

As a noticeable difference between the two interaction types, learners of the system group overall used fewer modifiers per total requests compared to learners from the SCMC group: While learners interacting with the dialog system used total 0.424 modifiers per total requests, learners interacting through SCMC used total 0.564 modifiers per total requests.

With respect to the individual modifier types, learners from the system group used fewer agreeers, grounders, alerters, and politeness markers per requests, but slightly more modal preterites and appreciators as well as more disagreeers. Overall, it can be noticed that learners from the system group tended to use proportionally fewer external modifiers (i.e., agreeer, grounder, alerter, disagreeer, appreciator) than internal modifiers (i.e., modal preterite, politeness marker) compared to learners from the SCMC group: Learners from the system group used total 61 external modifiers (i.e., combining agreeers, grounder, alerter, disagree, appreciator), which amounts to 53.0% of total 115 modifiers. In the SCMC group, 63 (71.6%) of the 88 learner modifiers were external modifiers.

Table 53
Frequency and Distribution of Subcodes for Modifiers of Requests in Learner-System Interaction and SCMC

Subcode	System			SCMC		
	Count	%	Per total requests (<i>n</i> = 271)	Count	%	Per total requests (<i>n</i> = 156)
Modal preterite ^a	47	40.9	0.173	20	22.7	0.128
Agreer ^b	26	22.6	0.096	36	40.9	0.231
Grounder ^b	12	10.4	0.044	15	17.1	0.096
Alerter ^b	8	7.0	0.030	8	9.1	0.051
Disagreer ^b	8	7.0	0.030	1	1.1	0.006
Appreciator ^b	7	6.1	0.026	3	3.4	0.019
Politeness marker ^a	7	6.1	0.026	5	5.7	0.032
Total	115	100.0	0.424	88	100.0	0.564

Note. Subcodes are ordered by highest frequency in learner-system interaction.

^a Internal modifier, for definition see section Coding of Requests. ^b External modifier, for definition see section Coding of Requests.

Learners' Perception on Accuracy

Accuracy was regularly mentioned by learners during the interviews. An overview of the derived codes with examples can be found in Appendix M. Learner statements about accuracy were mentioned in terms of, for example, correct or incorrect spelling, use of vocabulary, and correct or incorrect phrasing of longer units such as phrases or clauses. In a first instance, the learner statements could be classified according to their focus, i.e., the type of error that the statement focused on. It was found that learners from the system groups focused on spelling, lexis, or general accuracy. Learners from the SCMC group additionally focused on syntax.

Learners furthermore stated reasons for the accuracy or lack of accuracy, respectively, of their production (for a complete list of reasons with examples, see Appendix M). In both groups, learners mentioned in a few instances that they increased the accuracy to enhance understanding by the interlocutor. In some

instances, the learner did not specify any reason for enhancing the accuracy in their production.

When commenting on the reasons for the lack of accuracy, the learner statements included a wide range of reasons (for a complete list with examples, see Appendix M). Noticeably, the two groups also reported different kinds of reasons. Only four of the total 12 reasons were mentioned in both groups. Among these were that the learners lacked the knowledge, that work on accuracy would have impeded chatting speed, that they mistyped, and that the interaction was considered a low-stakes exercise.

Reasons that were exclusively mentioned by the learner-system group were that the learner assumed a lack of accuracy would not impact on understanding and that they had lost their patience to produce accurate language.

In the SCMC group, the mentioned reasons were more diverse. Learners argued that their production lacked accuracy because they assumed that the interlocutor would let them know if they were unable to understand something, that they lacked concentration, were cognitively overloaded, or experienced anxiety. A further reported reason was the 'relaxed' character of the chat modality. In one instance, the learner did not specify any reason for the lack of accuracy.

4.3.3 Communicative Adequacy

The descriptive statistics of the communicative adequacy measures along with the results of the statistical tests for differences between the two interaction types are presented in Table 54.

Since all three variables were ordinal-scaled, Mann-Whitney *U* tests were used to compare the measures between system interaction and SCMC.

The test on task completion barely reached significance with a *p*-value of exactly .050, indicating significantly lower task completion scores for learner-system interaction ($M = 1.86, SD = 0.48$) compared to SCMC ($M = 2.22, SD = 0.65$), corresponding to a medium effect size ($U = 131.00, p = .05, d = -0.63$). However, negligible non-significant differences between the two interaction types were observed

for the generic structure scores ($U = 175.00$, $p = .663$, $d = -0.14$) as well as generic style scores ($U = 195.00$, $p = .818$, $d = 0.08$).

Table 54

Descriptive Statistics and Results of Mann-Whitney U Test for Differences in Communicative Adequacy Measures Between Learner-System Interaction and SCMC

Measure	M (SD)		Mann-Whitney U	p	Cohen's d
	System (n = 21)	SCMC (n = 18)			
Task completion score	1.86 (0.48)	2.22 (0.65)	131.00	.050*	-0.63
Generic structure score	1.90 (0.70)	2.00 (0.69)	175.00	.663	-0.14
Generic style score	2.10 (0.44)	2.06 (0.54)	183.00	.818	0.08

* $p < .05$.

The descriptive overview in Table 55 provides additional insight into the differences in task completion ratings between the two interaction types. As can be seen, the majority of learners interacting with the dialog system were partly successful in completing the task, while one learner was successful. Almost a fifth of all learners interacting with the dialog system were unsuccessful in completing the task. The proportion of unsuccessful learners in learner-system interaction is higher than in SCMC (system: 19.0%; SCMC: 11.1%). Additionally, in SCMC, a third of the learners completed the task successfully, which is considerably higher than that found in learner-system interaction (SCMC: 33.3%; system: 4.8%).

With regard to the distributions of generic structure and generic style ratings, in both interaction types, most dialogs were structurally appropriate and showed a mixed style. In learner-system interaction, the proportion of inappropriate ratings on generic structure was higher than in SCMC (system: 28.6% vs. SCMC: 22.2%). Conversely, the proportions of partly genre-appropriate and fully genre-appropriate ratings were slightly lower in learner-system interaction compared to SCMC (system: 19.0% fully genre-appropriate, 52.4% partly genre-appropriate vs. SCMC: 22.2% fully genre-appropriate, 55.6% partly genre-appropriate).

With respect to the style ratings, the proportion of mixed style was slightly higher in learner-system interaction compared to SCMC (system: 81.0%; SCMC: 72.2%), whereas the proportion of formal style was slightly lower (system: 80.1%; SCMC: 72.2%). The proportion of informal style was less than half as high in system interaction as in SCMC (system: 4.8%; SCMC: 11.1%), although this finding is inconclusive due to the small number of overall counts.

Table 55

Frequency and Distribution of Subcodes/Levels of Communicative Adequacy in Learner-System Interaction and SCMC

Subcode / Level	System		SCMC	
	Count	%	Count	%
Task completion				
Successful	1	4.8	6	33.3
Partly successful	16	76.2	10	55.6
Unsuccessful	4	19.0	2	11.1
Generic structure				
Fully genre-appropriate	4	19.0	4	22.2
Partly genre-appropriate	11	52.4	10	55.6
Inappropriate	6	28.6	4	22.2
Generic style				
Formal	3	14.3	3	16.7
Mixed	17	81.0	13	72.2
Informal	1	4.8	2	11.1

Learners' Perception on Communicative Adequacy

The learners' feedback in the post-questionnaire and interviews provided additional insight into aspects relating to communicative adequacy, which in this study included the aspects task completion, generic structure, and generic style (see section Communicative Adequacy Measure in chapter Methods).

Figure 20 presents the learners' questionnaire responses on items with Likert scales relating to aspects of communicative adequacy, i.e., *I enjoyed the task; I thought the task was interesting; The task is useful for learning English; I made an effort; I completed the task successfully; My English competence was sufficient for the task; The*

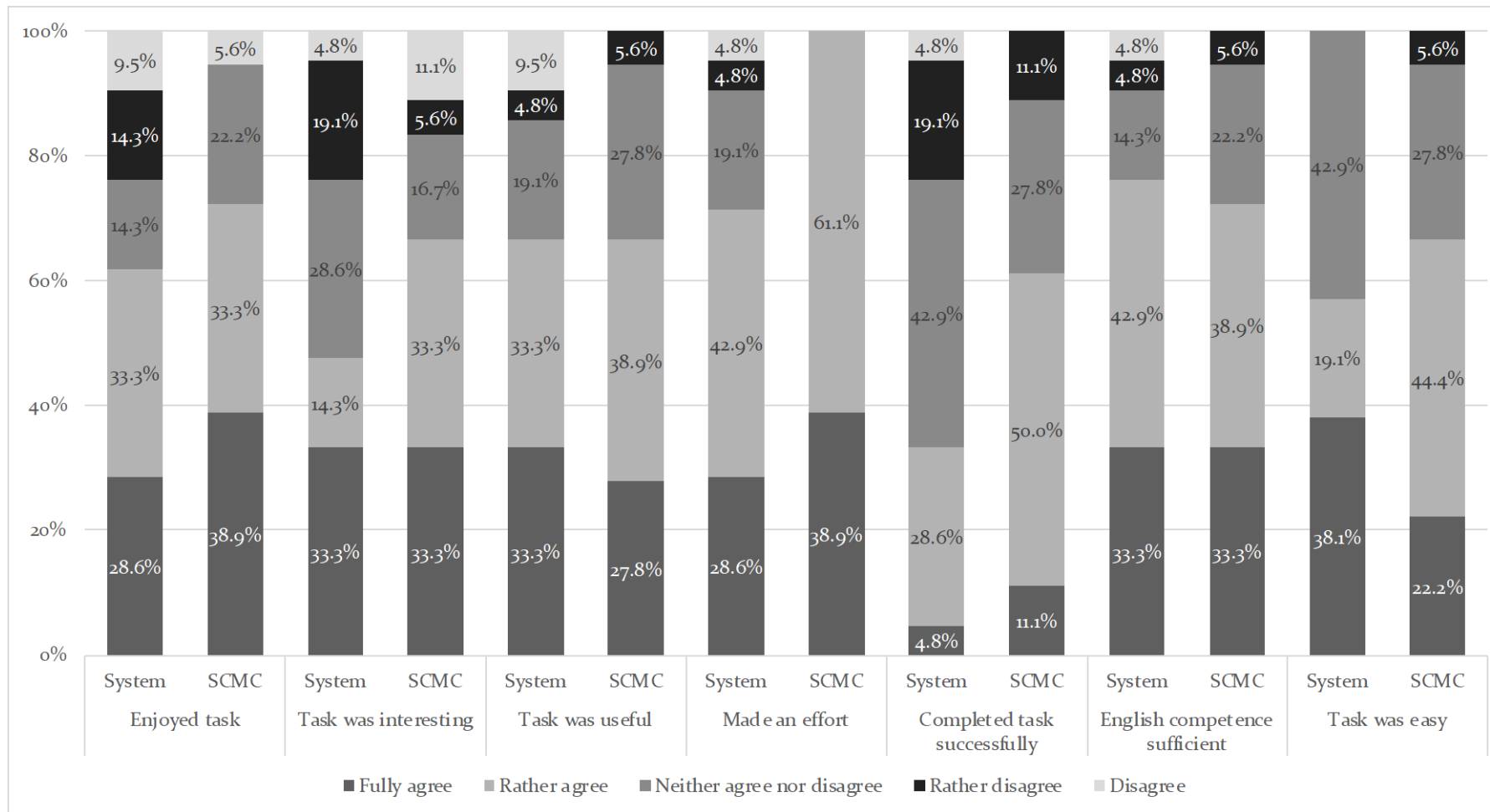
task was easy. As seen in Figure 20, learners in the system group tended to rate the task somewhat more negatively than the SCMC group: In the system group, a higher proportion of learners disagreed or rather disagreed that they enjoyed the task (38.1%) compared to the SCMC group (27.8%). They were also more ambivalent about the question whether they found the task interesting, with about half of the learners choosing an option on the lower end of the scale, i.e., *neither agree nor disagree*, *rather disagree*, or *disagree*, compared to the SCMC group where only about one third of learners chose one of the options at the lower end of the scale. Similarly, learners interacting with the dialog system indicated lower effort than those interacting with a peer: Whereas in the SCMC group all learners (100.0%) responded that they made an effort to complete the task, the proportion was lower in the system group (71.4%). However, both groups found the task similarly useful for learning English, with about two thirds of learners in both groups choosing *rather agree* or *fully agree*.

Learners from the two groups also agreed on the question whether their English skills were sufficient to complete the task: In both groups, about three quarters of respondents felt their English skills were sufficient, whereas the remaining quarter felt less confident and responded with *neither agree nor disagree* or *rather disagree*, while one learner from the system group even chose *disagree*.

Finally, respondents from the system group tended to find the task more difficult than the SCMC group, with 42.9% participants in the system group responding with *neither agree nor disagree* to the item *The task was easy*, compared to 27.8% in the SCMC group. One additional learner (5.6%) from the SCMC group responded with *rather disagree*. Similarly, a smaller proportion of learners from the system group felt that they completed the task successfully: While about two thirds of learners from the SCMC group fully or rather agreed on the item, the proportion was only one third in the system group.

Figure 20

Distribution of Questionnaire Responses to Questions on Aspects of Communicative Adequacy



Note. System: n = 21. SMC: n = 18.

The following findings relate to the learners' perceptions on generic style. In the questionnaire, learners were asked in three different items to select from a list of stylistic characteristics which marked their own production, which would have been suitable to use in the task, and which marked their interlocutor's production. The characteristics were presented in the form of polar opposites with one pole representing formal style and the other informal style (e.g., *oral style – written style, professional – private, Business English – General English*).

The findings revealed that learners from both groups described their own production with mostly informal characteristics: In the system group, learners selected 63.5% (corresponding to 120 characteristics) informal characteristics and 36.5% (69 characteristics) formal characteristics. In the SCMC group, the difference was less pronounced with learners selecting 55.6% (90 characteristics) informal and 44.4% (72 characteristics) formal characteristics. So, while both groups viewed their own production as rather informal, for the system group the informality of their own production seemed to be more pronounced.

When asked about the characteristics suitable for the task, learners in both groups selected mostly formal characteristics: In the system group, learners selected 76.7% (143 characteristics) formal and 24.3% (46 characteristics) informal characteristics. Learners from the SCMC group selected 69.1% (112 characteristics) formal and 30.9% (50 characteristics) informal characteristics.

However, the two groups diverged on the formality rating of the interlocutor's production: In the system group, learners selected 82.0% (155 characteristics) formal characteristics to describe the interlocutor's production and 18.0% (34 characteristics) informal characteristics. In contrast, learners from the SCMC group selected 45.7% (74 characteristics) formal and 53.73% (87 characteristics) informal characteristics. So, while in the system group, learners tended to rate their interlocutor's production as mostly formal, learners from the SCMC group tended to perceive their interlocutor's production as rather informal.

In an additional item on style, learners were asked whether they wrote as they would consider suitable for the task. Most learners from the system group reported that their own style diverged from the suitable style, with 76.2% (16 respondents) rating it as not suitable and 23.8% (five respondents) rating it as suitable. In the SCMC group, a weak majority of learners felt that they had written as was suitable for the task: 44.4% (eight respondents) considered their own production not suitable, while 55.6% (10 respondents) thought it was suitable.

In an additional open-format question learners gave several reasons for the discrepancy between the style that would have been suitable and the style of their own production. Most learners in both groups indicated that they lacked the language skills or that it would have been too much effort to produce a more formal style (system: six respondents or 28.6%; SCMC: three respondents or 16.7%). One learner in each group felt that the chosen style would be more suitable for chat or was unsure what style would be suitable for this kind of task (system: 23.8%; SCMC: 11.1%). One learner from the SCMC group (5.6%) responded that they wanted to accommodate their interlocutor by writing comprehensibly and sending the messages off as quickly as possible. Another learner from the SCMC group (5.6%) mentioned that they chose this style because the dialog was not “real”, but only a simulation. In the system group, two learners (9.5%) argued that they lost motivation to put effort into their writing style when the dialog system could not understand them.

In the interviews, learners also commented on aspects of communicative adequacy. With regard to task completion, learners commented on the strategies they used to solve the task as well as their perception of the task. A list of the codes categorizing the reported strategies along with examples from both the learner-system and SCMC data can be found in Appendix N. It was found that learners used a range of selection strategies to fulfill the task, i.e., select and book a suitable room: Learners from both groups requested the interlocutor’s assistance, chose to ask about the complete list of requirements one by one, or compared the information about the rooms against the

requirements before selecting. In the system group, learners additionally indicated that they selected a room before having collected all the required information or that they selected a room based on personal prioritization of the requirements. Learners from the SCMC group instead additionally mentioned that they reduced the selection by excluding rooms that appeared unsuitable or that they followed the interlocutor's suggestion. In one instance, a learner from the SCMC group indicated that they were unsure about what strategy to follow.

As an additional finding on task completion, learners had mixed impressions of the task (for a list of the learners' perceptions with examples, see Appendix N). In both groups, the learners' positive perceptions included that the task was fun, easy, and useful for learning. However, learners in both groups also indicated negative perceptions, such as that they thought the task's usefulness was limited, that it was difficult to solve due to communication issues, or because it was challenging to keep track of the conversation. Additionally, in the SCMC group, learners stated that they were uncertain if the interlocutor's information was reliable. Some learners from the SCMC additionally thought that the task was stressful, for example because they did not know which classmate they were chatting with or because they were worried about making mistakes.

The code matrix in Figure 21 provides additional insight into the system group's mixed perception of the task: As seen, all learners from the system group thought the task was, in principle, fun to do. It also needs to be noted that most learners enjoyed the task because of its novelty, as becomes clear in the examples for the subcodes *fun* and *useful* in Figure 21. Three of the learners (KM8-5TV, 357-KW8, 4XE-72U) had an otherwise predominantly negative perception of the task, commenting on its limited usefulness and the difficulty to keep track of the conversation and/or to communicate. On the other hand, two other learners (CW7-ZDE and 5CT-2N9) thought it was easy to solve and useful for learning English because they could apply what they had learned in class or because it allowed them to focus on the writing. One learner (RC5-SP9) had a mixed impression, thinking the task was fun and easy but doubted its usefulness for learning.

Figure 21

Code Matrix of Subcodes for Task Completion per Learner in Learner-System Interaction

Code System	CW7-ZDE	KM8-5TV	357-KW8	4XE-72U	5CT-2N9	RC5-SP9
COMMUNICATIVE ADEQUACY						
TASK COMPLETION	■	■	■	■	■	■
Perception of task						
Fun	■	■	■	■	■	■
Easy	■				■	■
Useful					■	
Usefulness limited				■		■
Difficult		■	■	■		

The following interview findings relate to the genre appropriateness of the learners' production, more precisely, the genre appropriateness of their productions' style and structure. With regard to style, the learners named several reasons for their conformity to and deviations from, respectively, a formal style. An overview of the reported reasons with examples can be found in the appendix (Appendix N). Learners from the system group reported that they wrote formally to appear professional, because they modeled their production after an email text, or for no specific reason. Learners from the SCMC group gave no specific reasons for using a formal style.

Relating to the deviations from formal style, learners from both groups indicated that they felt a less formal style would be appropriate, because they lacked the competence to produce formal style, or because they were unsure what style would be appropriate. In some instances, they did not further specify why they deviated from using a formal style. Additionally, learners from the system group chose a less formal style because of their interlocutor's irritating behavior. In the SCMC group, on the other hand, learners explained that they chose a less formal style because the interlocutor was a classmate rather than an actual hotel employee, because it corresponded to their usual chatting habit, or because they wanted to enhance understanding by writing less formally.

With respect to the structure of their production, learners commented on the strategies they used to structure their interaction. The strategies could be divided into two main groups: One set of strategies served to support the generic structure of the

interaction, while another set can be expected to lead to deviations from the generic structure. A complete list of codes pertaining to the learners' behaviors enhancing conformity to the generic structure or diminishing conformity to the generic structure, respectively, can be found in the appendix (Appendix N).

In both the system and SCMC group it was found that learners generally focused on ways to ensure the relevancy of the contributions to the conversation such as making sure to stay on topic, managing the interlocutor's deviations, or adapting to the interlocutor's preceding message to ensure the relevancy of the own contribution. Notably, only learners from the SCMC group stated that they waited for their interlocutor to adapt their messages to the interlocutor's.

Statements about behaviors that might have diminished the conformity to the generic structure in both groups included the lack of following a strategy or choosing to provide all information at once, i.e., in one single message at the beginning of the interaction. Another strategy diminishing the conformity to the generic structure was the learner being guided by the interlocutor, i.e., following the interlocutor's lead during the interaction.

It needs to be noted that following the interlocutor's lead might of course also assist the learner in following the generic structure, given the interlocutor guides the learner accordingly. In fact, one learner of each group felt that the interlocutor helped them structuring the interaction, as can be seen in the examples for this subcode. So, this strategy's role in enhancing or diminishing, respectively, the generic structure might be ambiguous and dependent on the interlocutor's specific behavior.

Two additional behaviors diminishing the generic structure were found exclusively in the system group: Learners interacting with the dialog system mentioned that they extended the conversation beyond the topic of the task because they wanted to test or try out the limits of the dialog system. One learner deviated from the generic structure as they had to 'restart' the conversation due to an interlocutor's non-understanding, as they stated in the interview.

4.3.4 Summary of Findings on Research Question 2

This subsection summarizes the findings on the comparison of learners' task performance in learner-system interaction and SCMC relating to research question 2. The findings will be discussed and interpreted in detail in the next chapter (see chapter Discussion).

Learners' task performance was assessed by a set of measures relating to linguistic complexity and accuracy as well as communicative adequacy which were subsequently tested for statistical differences between the two interaction types. With respect to the linguistic complexity measures, significant differences were found for two of the three measures. Thus, learners' task performance in learner-system interaction showed significantly lower lexical diversity as measured by MTLT than SCMC with a medium to large effect size (system: $M = 45.33$, $SD = 13.41$; $t(37) = -2.27$, $p = .029$, $d = -0.72$) and contained significantly fewer words per AS-unit with a large effect size (system: $M = 5.48$, $SD = 1.29$, SCMC: $M = 8.42$, $SD = 4.15$; $t(37) = -2.88$, $p = .009$, $d = -0.95$). The effect on syntactic similarity as measured by the SYNSTRUTt index was negligible and not significant (system: $M = 0.10$, $SD = 0.03$, SCMC: $M = 0.10$, $SD = 0.03$; $t(37) = 0.24$, $p = .812$, $d = 0.08$).

With respect to linguistic accuracy, learners from the system group produced more accurate language in terms of errors per AS-unit and error-free AS-units than those from the SCMC group, although the effects on both measures were small and non-significant. Thus, learner-system interaction showed fewer learner errors per AS-unit than SCMC (system: $M = 0.66$, $SD = 0.31$, SCMC: $M = 0.72$, $SD = 0.27$; $t(37) = -0.71$, $p = .482$, $d = -0.21$) and more error-free AS-units per AS-unit (system: $M = 0.61$, $SD = 0.13$, SCMC: $M = 0.56$, $SD = 0.31$; $t(37) = 1.14$, $p = .260$, $d = 0.37$).

However, with respect to the specific accuracy measures pertaining to the appropriateness of requests, learners interacting with the dialog system produced more inappropriate requests per total requests, although the effect was small and did not reach statistical significance (system: $M = 0.26$, $SD = 0.20$, SCMC: $M = 0.19$, $SD = 0.19$, $U = 155.00$, $p = .334$, $d = 0.34$). Learners' mean request appropriateness scores were also slightly lower

in learner-system interaction than in SCMC, but this effect was negligible in size and non-significant (system: $M = 1.96$, $SD = 0.40$, SCMC: $M = 2.01$, $SD = 0.29$, $U = 160.50$, $p = .422$, $d = -0.13$).

Descriptive findings on error types revealed that learner-system interaction led to a somewhat higher proportion of grammatical errors and lower proportion of spelling errors compared to SCMC. Looking at request appropriateness, learner-system interaction compared to SCMC led to a noticeably higher proportion of inappropriate requests and a lower proportion of moderately appropriate requests.

Of the three communicative adequacy measures, task completion was significantly lower in learner-system interaction) compared to SCMC, with a medium effect size (system: $M = 1.86$, $SD = 0.48$, SCMC: $M = 2.22$, $SD = 0.65$; $U = 131.00$, $p = .05$, $d = -0.63$). However, the interaction type had negligible and non-significant effects on both the generic structure score as well as the generic style score of the learners' performances.

A descriptive overview revealed relatively more ratings at the negative end of the *task completion* and *generic structure* rating scales for learner-system interaction and more ratings at the positive end of the scales for SCMC. Thus, a higher proportion of learners interacting with the dialog system were unsuccessful in completing the task, while a higher proportion of learners interacting through SCMC were successful. Similarly, a higher proportion of learners' task performances from the system group was rated as inappropriate, while a higher proportion of performances from the SCMC group were rated as fully genre-appropriate, although, as noted above, the different distributions did not lead to a significant effect in the overall appropriateness rating between the two interaction types.

The following paragraphs sum up the findings on the learners' perception on communicative adequacy. When asked about the communicative adequacy aspect *task completion* in the questionnaire, learner responses showed that learners from the system group tended to rate the task somewhat more negatively in terms of task enjoyment and task interest compared to those from the SCMC group, and they also indicated lower engagement/effort. However, the two groups equally considered the task useful for

learning English. Furthermore, while learners from both groups considered their English skills sufficient to solve the task, learners from the system group tended to rate the task as less easy than the SCMC group and less frequently felt that they completed the task successfully. Further findings from the interviews provided additional details on the learners' mixed impressions of the task, with learners from both groups confirming that the task was, in principle, fun, easy, and useful for learning. Two of the six learners had a completely positive impression of the task, perceiving it as fun, easy, and helpful. However, most learners from both groups also mentioned negative aspects such as that the task's usefulness was limited and that it was difficult to solve due to communication issues and to keep track of the conversation. Additionally, learners from the SCMC group stated doubts about the reliability of the interlocutor's information and some found the task socially stressful. The interview findings on task completion further revealed that learners from both groups used some of the same strategies to complete the task (e.g., requested the interlocutor's assistance, chose to ask about the complete list of requirements one by one) but also chose interaction-specific strategies such as selecting a room before having collected all the required information or selecting a room based on personal prioritization of the requirements (system) or reducing the selection by excluding rooms that appeared unsuitable or following the interlocutor's suggestion (SCMC).

With regard to the aspect *generic style*, learners from both groups in the questionnaire described their own production with predominantly informal characteristics. The perception of informality of their own production was more pronounced for learners interacting with the dialog system. The two groups also diverged in their perception of the interlocutor's use of style, with learners from the system group clearly rating the dialog system's production as formal and learners from the SCMC group rating their interlocutor's production as rather informal. Furthermore, while in the system group a clear majority of the learners stated that their own style diverged from the style that would be suitable for the task, a small majority learners from the SCMC group felt that their own style was suitable. As reasons for the discrepancy, learners in both

groups indicated lack of language skills or that a formal style would have required too much effort, that the chosen style seemed suitable for the chat context, and uncertainty about what style would be suitable. There were also a few mentions specific to the interaction type, such as wanting to accommodate the interlocutor, wanting to send the messages off as quickly as possible, and that it was only a simulation rather than a real dialogue (SCMC), and lack of motivation due to non-understanding by interlocutor (system). The interview findings on generic style reinstated some of the same reasons for the same groups such as deviating from a formal style such as lack of competence, considering a less formal style as appropriate, and being unsure about appropriate style (both groups), accommodating the interlocutor and that the interlocutor was not real (SCMC), the interlocutor's irritating behavior (system). For the SCMC group, the interviewed learners additionally introduced personal chatting habits as a reason for using a less formal style.

Interview findings also contained statements about the structure of their production. Learners from both groups mentioned using both strategies that enhanced conformity to the generic structure and strategies that diminished conformity to the generic structure. Some strategies were mentioned by learners from both groups. Thus, both groups used strategies such as staying on topic and managing interlocutor deviations to support the generic structure. Also mentioned in both groups were strategies reducing the generic structure, such as following no specific strategy, providing all information at once, or being guided by the interlocutor. However, the structure-supporting strategy of waiting for the interlocutor to adapt own messages to those of the interlocutor was only mentioned by learners from the SCMC. On the other hand, only learners from the system group reported to use the structure-diminishing strategies extending the conversation beyond the required topics and restarting the conversation from the beginning.

5 Discussion

This chapter integrates and jointly interprets the findings from the different data sources (interactions, questionnaires, and interviews) presented in the previous chapter and discusses them in terms of the aims and research questions of this study. As presented above (see chapter Introduction), a first aim of this study was to contribute to the understanding of dialog system interaction in the context of EFL learning by describing the interactional features and learners' task performance in task-based learner-system interaction. A second aim was to inform EFL teachers on the potential benefits and challenges as well as practical implications of using dialog systems in their language classrooms.

In meeting these aims, the study addressed the following two research questions:

1. *What interactional features occur in written task-based interaction between commercial VET EFL learners and a dialog system and what are their characteristics compared to those found in learner-peer written SCMC on the same task?*
2. *Is there a difference in the learners' task performance in terms of linguistic complexity and accuracy as well as communicative adequacy between learner-system interaction and learner-peer SCMC?*

The present chapter first provides a summary and interpretation of the major findings for each research question, integrating the findings from the different data sources as well as the findings from previous studies and discussing them in light of their implications for FL learning research and teaching.

5.1 General Overview of Interactions

As an initial observation on the general measures of the interactions (i.e., task on time and text amount), learner-system interaction produced a significantly larger amount of text in significantly less time on task compared to SCMC. When interacting with the dialog system, learners produced both more total messages and more words than those

interacting with a peer through SCMC. They also received more interlocutor input as measured by total messages and words produced by the interlocutor. This may partly be explained by the frequent indications of non-understanding in learner-system interaction and the interlocutors' need to rephrase and restate messages. Learner-system interaction's higher efficiency in terms of producing more input and output in less time may be explained by the dialog system's answering speed which was faster compared to that of the peer interlocutors. The relatively slow answering speed in SCMC not only limited the amount of text learners and peers produced within the given time, but also impacted on the learners' engagement and motivation.

5.2 Research Question 1

In the present study, learner-system interaction was found to generate the three interactional features *negotiation of meaning*, *recasts*, and *lexical and structural alignment*. The following subsection will restate and discuss the major findings on the characteristics of each interactional feature found in learner-system interaction compared to those found in SCMC.

5.2.1 Negotiation of Meaning

As an initial key observation on negotiation of meaning, learner-system interaction led to a high number of negotiation routines. The number of indicators of non-understanding in system interaction was a total of 218 (10.38 per interaction; 0.16 per message). This observation confirms the finding from the two other studies on negotiation of meaning in learner-system interaction by Kim (2017b) and Satar and Qinghua (2020) that this interaction type provides ample opportunity for learners to engage in negotiation of meaning. Comparing the number of negotiation sequences found in the present study to previous studies, it was similar to the number observed by Kim (2017b) who found from $M = 13.85$ to $M = 16.7$ negotiation episodes per 10-minute interaction, depending on learner proficiency and whether the interaction was at the beginning or end of a 10-week intervention. However, the negotiation rate was higher in

the current study compared to Satar and Qinghua (2020) who reported for 30-minute interactions with two different dialog systems $M = 3.25$ and $M = 5.25$ negotiation episodes per interaction, respectively. It can be expected that the number of negotiation episodes would be even lower if the interactions had been shortened to the same length as the interactions in the present study. The difference in numbers is likely a result of factors such as the specific system used in the different studies as well as learner characteristics, such as different levels of proficiency and previous experience chatting with dialog systems (N.-Y. Kim, 2017b; Satar & Qinghua, 2020). This observation shows that more research is needed to investigate the factors impacting negotiation of meaning in learner-system interaction and that the generalizability of the findings on a specific system and context to other systems and contexts have to be further demonstrated.

Comparing the number of negotiation sequences between the two interaction types in this study further highlights the considerable potential of learner-system interaction in providing learners with negotiation opportunity. In the current study, SCMC led to a total of five negotiation routines across all 18 interactions (0.28 per interaction; 0.01 per message). Previous studies found similarly low numbers of negotiation of meaning in written SCMC. For example, Fernández-García and Martínez Arbeláiz (2003) found in 20-minute written SCMC interactions a total of nine negotiation routines in 14 learner-learner dyads and a total of four negotiation routines in four learner-native speaker dyads. Similarly, Loewen and Wolff (2016) found between $M = 2.51$ and $M = 6.4$ negotiation routines per interaction on three different 20-minute written SCMC interactions. A possible explanation for the low frequency of negotiation of meaning in written SCMC offered by both studies is that the written modality supports the learners' mutual understanding, reducing the need for negotiation of meaning. In the present study, learners from the SCMC group confirmed in the questionnaire that mutual understanding was high, and that if non-understanding did occur, they chose non-interactive means such as checking in the chat over interactive means to deal with it, leading to the low numbers of negotiation of meaning in the SCMC interaction type. In the learner-system group, on the other hand, mutual understanding was severely

impacted by the dialog system's limited ability to classify the learners' messages correctly. As documented above (see section Dialog System), the dialog system used in this study classified 70% of user messages correctly across three usability tests. It must be specified that this shortcoming primarily reduced the dialog system's 'understanding' of the learner but not the learner's of the system, which learners also observed in their questionnaire responses. This also explains the finding that negotiation of meaning in learner-system interaction was almost always initiated by the dialog system instead of the learner (214 initiated by dialog system vs. four by learners).

So, while learner-system interaction in this study did lead to a high number of negotiation episodes, their efficacy in facilitating FL learning needs to be further qualified by the additional findings on its specific characteristics. Thus, the above-mentioned finding that learner-system negotiation was almost always initiated by the dialog system rather than the learner suggests that their function needs to be discussed primarily in terms of providing feedback and eliciting modified output: by indicating non-understanding, the interlocutor provides the learner with implicit feedback on their production, prompting them to produce modified output. On the other hand, the feature's function in generating interactionally modified input making it more comprehensible to the learner, can be considered negligible in the learner-system interaction in the present study. This is especially true as in three out of four learner-initiated episodes of negotiation of meaning, the system abandoned the topic, thus failing to produce more comprehensible input to the learner.

With respect to the feedback-function of the negotiation of meaning found in the system interactions, three further points need to be noted: First, target-like learner utterances were almost as likely to trigger an indicator by the dialog system as were erroneous utterances (i.e., utterances containing linguistic and/or discourse errors). As opposed to a human interlocutor whose understanding is mainly impeded by linguistic or discourse errors, the dialog system's 'understanding' was influenced by a complex interplay of multiple factors that were undisclosed to the human user, thus leading to a pattern of non-understanding that might seem inconsistent, unsystematic, 'random' to

learners: Linguistically immaculate utterances may trigger an indicator (= false positive), whereas erroneous utterances may not (= false negative). In other words, indicators of non-understanding uttered by the system were no reliable indication to the learner that their utterance was linguistically or discoursally inaccurate. As a consequence, learners had to primarily rely on monitoring in evaluating whether or not their utterance was in fact incorrect.

Second, system-produced indicators generally did not point to the specific item or aspect in the trigger utterance that caused non-understanding. About half of the indicators of non-understanding were realized as clarification requests, generally indicating global non-understanding (e.g., “Excuse me, can you please rephrase your last message?”) and the other half as inappropriate responses. Only a few indicators were realized as confirmation checks, asking the learner to confirm a booking date they had already stated in their preceding message. Methodologically, this prevented a classification of the trigger utterances. From a language learning point of view, this meant that learners generally had no indication to attribute the interlocutor’s non-understanding to a specific lexical item, grammatical construction, or discourse error in their trigger utterance. As a consequence, learners had to rely on their own interpretations of what might have caused the non-understanding. This became evident in the interview data where learners named a number of conceptions to account for the non-understanding by the dialog system. Interestingly, some of the learners’ conceptions about the dialog system’s processing were true, despite the learners’ little prior experience interacting with dialog systems. As the results further showed, a certain conception, e.g., “the dialog system cannot understand all words”, generally seemed to be met with a related strategy, e.g., rephrasing of the trigger message. This indicates that accurate conceptions and thus the adequate use of strategies might lead learners to managing the interaction in a way that resolves non-understanding. However, several learner conceptions were only partly accurate. In addition, learners frequently stated that they assumed the fault was theirs but had no clear understanding on what aspect of their production might have led to non-understanding. This indicates that learners were

unclear as to what was going wrong in their interaction. Both inadequately seeing the fault in their own production as well as having partly accurate conceptions might be problematic. Attributing the non-understanding to themselves instead of the interlocutor represents a type of false negative and inhibits accurate self-monitoring. And, as could be seen when relating conceptions to strategies, inaccurate conceptions might lead to inadequate strategies and from there to continued non-understanding. A pedagogical implication of this is that learners could benefit from an introduction in the functioning of a dialog system.

Third, the high rate of indications of non-understanding in learner-system interaction seemed to have caused various negative emotional reactions in the learners. The interview data from the system group revealed that all learners experienced some negative emotions in reaction to the system's non-understanding, such as frustration/annoyance, losing patience, or confusion. Negative emotions were also more frequently expressed compared to positive emotions such as curiosity or amusement, of which, in addition, the latter may be interpreted ambiguously. So, while the high number of indicators of non-understanding presented learners with frequent feedback and opportunities to modify their output, the drawback was that the associated negative reactions might have negatively impacted the learners' motivation to complete the task and to produce adequate language, as will be seen further down regarding complexity and accuracy as well as task completion.

With regard to the first two findings above on the observed indicator characteristics, it needs to be noted that they were not different, per se, from those found in human interaction, including the SCMC data of the present study as well as data from the literature, at least at the surface level. A first indication for the similarity can be found in the fact that the indicator sub-codes applied to the learner-system data in the present study were adopted from the literature on human-human interaction and could be applied to the SCMC data of the present study, too.

Thus, for instance, indicators in the SCMC data similarly followed target-like as well as erroneous utterances. However, examples from human-human literature on

negotiation of meaning make it clear that indications of non-understanding typically involve instances where learners lack the linguistic resources to understand or accurately produce an utterance. While negotiation of meaning following utterances that are linguistically error-free can be found in human-human interaction, for instance because learners lack a shared background with the interlocutor (e.g., in learner-native speaker interaction) or lack a shared goal (Gass & Varonis, 1985), negotiation of meaning following linguistically error-free triggers is much rarer than in learner-system interaction of the present study.

Similarly, human-human negotiation of meaning may also fail to identify the specific element that caused non-understanding, as the system-indicators in the present study did. However, while the dialog system in the present study never identified the problematic element, human interlocutors often do. Smith (2003), for example, in a study on learner-peer SCMC observed that *global* indicators not identifying the trigger accounted for merely 18% of all indicators.

Moreover, while the types of indicators found in learner-system interaction in the present study (i.e., clarification request, confirmation check, appropriate response) in general coincided with those found in the SCMC data as well as FTF interaction documented in the literature, there were some noticeable differences in the ways they were realized. Thus, clarification requests in learner-system interaction of the present study were linguistically uniformly realized as “Excuse me, can you please rephrase your last message?” or a variation thereof and confirmation checks solely involved the confirmation of booking dates, while a human would indicate non-understanding much more diversely. System indications also involved a very high proportion of inappropriate responses, which are rare in human interaction.

As a further key finding on negotiation of meaning in learner-system interaction, learners reacted to indicators of non-understanding in a variety of ways: A first type of reaction to an indicator was to exit the negotiation routine via a ‘non-response’, i.e., meta-comment, change topic, accept topic, or no response. This finding corresponds to

Yin's (2020) finding on learner-dialogue system interaction in which 14% to 50% of all negotiation of meaning routines ended before the response phase. In the present study, the data from the questionnaire and interviews revealed that learners in the system group consciously used a 'non-response' strategy, for instance, because they were annoyed with the dialog system or felt that the system's utterance was irrelevant for the further course of the interaction. So, learners interacting with the dialog system appeared to evaluate whether a non-understanding was in fact relevant to be resolved or could be ignored.

The rate with which learners ignore the dialog system's indications of non-understanding seems to depend on the learners' proficiency at least partly. Thus, Yin (2020) observed that the rate of missing responses was higher in high-proficiency learners than in low-proficiency learners. One explanation for this observation could be that proficient learners being more competent at monitoring their own production are more confident in identifying indicators that signal an actual problematic utterance that requires modification. So, 'false' indications of non-understanding by the dialog system might be less detrimental to proficient learners as they are better capable of distinguishing them from 'real' indicators.

While the same 'non-response' strategies were observed in the SCMC data, due to the low numbers it could not be established whether they followed the same patterns as in learner-system interaction. As an initial observed difference, the 'non-response' type *meta-comment* in SCMC successfully led to a resolution of the non-understanding, whereas it generally did not in learner-system interaction. This observation could indicate that the strategic competence, i.e., the competence learners need to manage communication breakdowns and facilitate effective communication (S. Sauro, 2011), that learners need in order to manage the interaction with a dialog system, might differ from that required in human-human interaction. For instances, while meta-language might be an effective tool in human-human interaction, it was ineffective in the learner-system interaction in the present study. From a language learning perspective, this would mean, on the one hand, that the transferability of strategic competence from learner-system to human-human interaction and vice versa could be limited and, on the other hand, that

strategic competence for learner-system interaction might need to be specifically trained if learners are to effectively use dialog systems for language learning.

As further reactions to indications of non-understanding, learners responded by (rarely) repeating or by rephrasing their original trigger utterance through means of elaboration, reduction, or modification. Since in the very few episodes of negotiation of meaning in SCMC only one response occurred, no conclusive comparison of the response characteristics in the two interaction types could be made. The response types above that were observed in learner-system interaction coincide with those found in human-human negotiation (e.g., Smith, 2003) and all represent, from a theoretical point of view, effective ways to produce modified output and thus facilitate language learning by enhancing fluency and accuracy, noticing, and hypothesis-testing (see section Output in chapter Literature Review).

With respect to the modified output produced in the responses, repair was inconsistent. Thus, less than half of the learners' transformations led to improved accuracy compared to the trigger utterance, while about a third led to equivalent and another third to reduced accuracy. As a note, accuracy was rated regardless of the severity of the errors committed and whether the original error remained or a new error was added.

Moreover, learners did not systematically receive reliable feedback on the accuracy and the improved understandability of their modified output. Depending on the transformation type, the dialog system 'understood' the learner's modified output in 40% – 60% of the cases. As explained above, the dialog system may, on the one hand, still not have understood a modified message even though its understandability – if judged by a human interlocutor – and accuracy had in fact improved (= false negative). On the other hand, the dialog system may have given positive feedback, i.e., understood a modified message, although its understandability and accuracy had not been improved (= false positive). It will need to be investigated further whether the limited success of repair and feedback in learner-system negotiation of meaning mitigates its benefits for the learners' accuracy development.

As a further observation, learner responses of the *elaboration* type were most successful in resolving non-understanding, followed by *reduction* and *modification*, and, lastly, *repetition*. Again, it will need to be examined further if this leads to learners prioritizing one over the other response strategy in order to enhance understanding, when interacting more frequently with a dialog system.

Another characteristic of learner-system negotiation of meaning were multi-layered negotiation routines, with the response functioning as a new trigger which set off another response or 'non-response', respectively. In SCMC, in comparison, no multi-layered routines occurred. This indicates that learner-peer negotiation of meaning successfully resolved the non-understanding at the first attempt whereas learners negotiating with the dialog system often needed additional trials.

The advantage of learner-peer over learner-system negotiation of meaning in resolving non-understanding was due to the fact that the dialog system's ability to 'understand' and produce language was limited by the training data it was provided (see section Dialog System): First, the dialog system could 'understand' learner utterances only if they were structurally close enough to the samples provided during training and could be matched to one of the limited set of pre-defined intents by drawing on the linguistic data of the utterance itself, as well as by inferring from the context of the conversation, i.e., a history of the previous entities, intents, and responses. Second, to respond, the dialog system selected from pre-formulated templates that which was appropriate to the intent and context it had determined.

These limitations in the dialog system's language processing and production led to failed or inefficient negotiation of meaning in learner-system interaction in different ways: For instance, the dialog system may have failed to assign the learner's response to the correct intent repeatedly in a given negotiation sequence, despite the learner's reformulations or it may have been unable to process the learner's meta-comments or other learner input, such as phatic discourse that had not been included in the training data. In addition, dialog system indicators and reactions to responses were static and not

adapted to the specific context. Humans on the other hand, have a much larger context to draw on to infer meaning and can produce language much more flexibly and creatively according to the specific context, which greatly facilitated the efficiency of the meaning negotiation. So, while previous studies on negotiation of meaning in learner-peer written SCMC (as well as in FTF interaction) also observed continued non-understanding and multi-layered negotiation of meaning sequences (Smith, 2003), learners abandoned the routine only “[o]ccasionally” (p. 51), which again points to the overall advantage of human-human over human-system negotiation of meaning in terms of successfully resolving non-understanding.

As an additional note, human-human interlocutors seemed to not only be more flexible, creative, and efficient when negotiating meaning but also more humorous, as, for example, seen in one learner’s reaction to response where they imitate the sound of a computer peep (“pipipi its booked”) to indicate that the non-understanding had been successfully resolved.

However, from a language learning theoretical standpoint, it can be argued that multi-layered routines may in fact allow learners to try out multiple ways of modifying the same output, thus testing different hypotheses on the well-formedness of their utterances. Moreover, multi-layered routines may reveal the learners’ persistence in trying to resolve non-understanding even after they already tried to do so at least once. Conversely, however, it could be seen that in multi-layered negotiation episodes learners tended to ultimately abandon the routine without resolving the non-understanding. In fact, the more layers, the less likely the routine was to be resolved. This may be linked to decreased learner motivation to continue modifying their output. As noted above, the learners’ interview data indicated that frequent non-understanding typically resulted in negative emotions such as frustration/annoyance and losing patience, which in turn might have caused learners to give up trying.

Finally, in learner-system interaction no reaction to response component as previously defined in human-human interaction (i.e., “okay”, “I understand now”, in the

present study coded as *explicit resolution*) could be observed. Instead, responses in learner-system negotiation of meaning were reacted to by either an implicit indication that the non-understanding had been resolved or an explicit or implicit indication that it had not been resolved. In SCMC, on the other hand, systematic reaction to response per the previous definition occurred, i.e., explicitly indicating understanding. The lack of a system-generated reaction to response can be explained with the fact that it was not incorporated in the system responses when developing the dialog system. On the part of the learners, they might have avoided it because they considered phatic discourse unnecessary or inadequate for a conversation with a non-human interlocutor. Another explanation could be that, in the preceding course of interaction, learners may have experienced that meta-language was typically not understood by the system and therefore tended to refrain from using it. For instance, as mentioned above, the response type *meta-comment*, responding to an indicator on a metalinguistic level, was generally not understood by the dialog system. The absence of explicit forms of reaction to response in learner-system interaction adds to the above-mentioned tendency for learners interacting with the dialog system to have to rely on monitoring rather than feedback from their interlocutor to assess the adequacy of their language.

To summarize the discussion on negotiation of meaning, learner-system interaction led to a high amount of, specifically, system-initiated negotiation of meaning, thus providing learners with frequent feedback and opportunities to produce modified output. Due to the unspecific nature of the system indicators, learners were required to use monitoring and hypotheses-building to resolve non-understanding and generate repair. However, hypothesis verification and adequacy judgement of the repair might have been aggravated by the learners' lack of understanding of the dialog system's functioning as well as the system's unreliable feedback on their responses. Moreover, frequent indications of non-understanding by the dialog system might have negatively impacted the learners' motivation to consistently complete the negotiation sequences. Furthermore, learners appeared to evaluate the relevance of each given indicator and

accordingly select the adequate strategy to deal with the non-understanding based on their conception of what might have caused then non-understanding. Learners subsequently used a variety of linguistic modifications, comparable to those found in human-human interaction, to respond to the indicator of non-understanding which could be beneficial for language development. Characteristics specific to learner-system negotiation of meaning were frequent multi-layered negotiation routines, which highlight both the high density as well as the inefficiency of learner-system negotiation of meaning in resolving non-understanding. Inefficient negotiation of meaning in learner-system interaction was caused by the dialog system's limited language processing and production abilities, leading to repeatedly inadequate interpretation of learner responses, as well as indicators and reactions to responses that were often not adapted to the specific context of the negotiation routine. A second interaction-specific characteristic was the absence of an explicit indication that the non-understanding had been resolved in the *reaction to response* move.

5.2.2 Recasts

As a further interactional feature found in learner-system interaction, the dialog system provided learners with implicit feedback in the form of recasts. In the present study, the occurrence of recasts was higher in learner-system interaction compared to SCMC. Without correcting for text amount, the number of recasts was about three times higher in system interaction than in SCMC. Per message, the incidence was still twice as high for system interaction compared to SCMC. So, learner-system interaction was more efficient in the production of recasts compared to SCMC by producing more recasts in the same amount of text. The higher prevalence of recasts in learner-system interaction compared to SCMC might have been due to fact that the system-generated language was free of errors and thus more likely to produce target-like in recast episodes than the messages produced by the learners' peers in SCMC.

System recasts were found to provide feedback on all error types, i.e., grammar, lexis, spelling, or multi-focus with most recasts focusing on grammar, followed by an

equal number of spelling and lexis recasts and two multi-focus recasts. The system recasts addressed a range of errors relevant for the participants' language learning, such as omission of prepositions or articles, incorrect inflection of nouns and verbs, and the semantically inappropriate choice of lexical items. In comparison, SCMC recasts lacked the lexical error type, although previous studies have observed lexical recasts in learner-peer SCMC (S. Sotillo, 2005). Grammar recasts were equally dominant in SCMC as in learner-system interaction, followed by spelling recasts, and multi-focus recasts. The specific errors observed in each linguistic category were also similar in the two interaction types. However, one marked difference between the two interaction types was that system grammar recasts frequently involved the superfluous use of the definite article *the* with room names (e.g., "the room Beta"), whereas SCMC recasts never focused on this specific error although it was also frequent in this interaction type, too. This indicates that system recasts might focus on errors that learner recasts exclude, for instance because a peer might be unaware that the given form is erroneous and reproduce the same error from the learner's production.

The further characteristics of system recasts can be discussed in terms of their impact on recast salience. Recast salience has been related to the learner's noticing of their own error and the target-like form which in turn facilitates the benefits of recasts for L2 and FL development (see section Feedback in chapter Literature Review). As a first observation, almost half of the items that were recast in learner-system interaction were recast another time in a separate recast sequence of the same interaction. Similarly, in a few instances, the recast form was repeated within the same recast move. By comparison, no repeated recasts within the same interaction occurred in SCMC, but there were a few instances where the recast form was repeated in the same recast move. One possible explanation for the higher prevalence of repeated recasts in learner-system interaction compared to SCMC might be the overall higher density of recasts in learner-system interaction increasing the chance for repeated incidences. As a second explanation, learner-system interaction often contained repetitions or variations of the same message as it frequently drew from a limited set of responses and negotiating meaning. It can be

argued that these repeated incidences increase the salience of a recast and thus the likelihood to be noticed by the learner.

Conversely, the salience of system recasts might have been mitigated by the extensive length of the recast move in which the target-like form was embedded. As previous studies on recasts have suggested, learners are more likely to notice a target-like form of their own erroneous production if it is isolated rather than embedded in an entire utterance (e.g., S. Sauro, 2009). In the present study, almost half of the system recast moves consisted of single sentences, while the other half was split almost equally between multi-sentence and multi-message moves. In SCMC, most recast moves consisted of multiple messages, whereas about a fifth were multiple sentences and one recast move consisted of a single sentence. So, while system recasts were somewhat advantageous over SCMC recasts by generally embedding the target-like form in single- or multi-sentence-moves rather than multiple messages, the recast moves in both interaction types might have generally been too long for learners to benefit from the feature to the full extent.

Another aspect observed in the data that may affect the salience and noticing of a recast is the recasts' completeness, i.e., whether a recast addressed all errors in the learner's preceding messages or only one of the errors. It was found that almost half of the system recasts were complete, whereas in SCMC no complete recasts occurred, in other words, all SCMC recasts left one or more errors unaddressed. From a theoretical perspective, it can be argued that recasts focusing on multiple errors at the same time might render each individual error-feedback less salient to the learner as the learner's attention is split on multiple errors. On the other hand, it can be argued for the benefit of complete recasts, in that they more 'thoroughly' provide target-like forms for all errors, whereas partial recasts leave one or more errors unaddressed. With respect to the complete system recasts, it needs to be pointed out that only one of them in fact involved more than one error, so the recasts provided 'thorough' feedback while not impeding the salience of the target-like forms by a multi-error-focus. While no studies known to the author have investigated learners' noticing of partial vs. complete recasts, it has been previously found that recasts involving multiple corrective changes to the trigger were

less well noticed by learners than those involving single changes, possibly because processing single changes was less taxing on the learners' working memory (Egi, 2007). It will need to be investigated further whether this observation from human-human FTF interaction on the advantage of recast forms with fewer changes might also apply to recast moves with fewer recast items in the written learner-system interaction context.

As a first observation on the uptake following recasts in the present study, uptake never involved an explicit acknowledgment of the feedback (e.g., "ok!") as previously observed in the literature. This is not surprising since all system recasts (as well as SCMC recasts) were highly implicit with the recast form embedded both semantically within the conversation context and syntactically within entire sentences or even multiple messages. As a consequence, learners might have chosen uptake that was similarly implicit.

Furthermore, uptake could be observed in half of the system recasts with the majority occurring not in the turn immediately following the recast move, but in a later turn as delayed uptake. The uptake rate in SCMC was lower with only two (corresponding to about a fifth) of the recasts being followed by uptake, one immediately and one delayed.

In addition, 40% of uptake in learner-system interaction consisted of the repaired form, the remaining 60% were still in need of repair. In SCMC, one of the uptakes contained the repaired form, the other one involved a form that was still in need of repair. As shown above, due to their shorter length and higher intensity recasts, system recasts might have been more salient compared to SCMC recasts, which could partly explain the relatively higher uptake and repair rate in learner-system interaction (Sheen, 2006), although the SCMC numbers were too low to conclusively compare the uptake and repair rate between the two interaction types. The overall rather low uptake and repair rate in both interaction types could have been caused by the fact that the task had a focus on meaning with focus on form being secondary and implicit which might have discouraged learners from actively engaging in feedback routines as long as the errors did not interfere with understanding (Sheen, 2004). However, while the uptake rate could have been higher, learners interacting with the dialog system reused the target-like form in a

considerable number of instances. Drawing on the findings from the learner interviews and questionnaires, the learners seemed to have worked on the correctness of their production, including by consciously drawing on their interlocutor's feedback, in an effort to enhance understanding and for further unspecified reasons. Possible implicit drivers might have been the school setting and the task requirement to write adequately to the context.

As an additional observation on uptake and repair in system recasts, most immediate uptake contained the repaired form whereas most delayed uptake was still in need of repair. This might be an indication that noticing effects might be short-term and decrease as the distance to the recast form increases, even in the written chat context where the recast form remains visible on the screen for some time.

Finally, a considerable amount of re-uptake was observed in learner-system interaction. In fact, more than half of learner uptake in learner-system interaction was followed by re-uptake, while there was no re-uptake in SCMC. Re-uptake in learner-system interaction might have been a result of learners often reusing lexical items and grammatical forms in the context of negotiation of meaning, which was frequent in learner-system interaction (see above) and might have increased the possibility of reusing a recast item.

To relate these observations on uptake and repair to the efficiency of the recasts in terms of language learning, uptake and repair are understood to be indicative of learner noticing, as explained above (see section Feedback in chapter Literature Review). Furthermore, uptake in language learning has to be considered in terms of its function in facilitating output. Output, in turn, plays an important role in automatization and the development of fluency and might be more effective when it involves learners drawing on their own knowledge rather than repeating the target form, as this pushes the learner to process at a deeper level (Li & Vuono, 2019). Along these lines, system recasts could be considered more beneficial than SCMC recasts, as they led to more uptake and even re-uptake and hence noticing and modified output. However, this proposition as well as the question how system recasts and their potentially differential effects on noticing and

modified output relate to the actual development of FL forms would have to be further tested empirically.

In sum, learner-system interaction provided learners with a considerable amount of implicit feedback in the form of recasts. Compared to the SCMC interaction type, learner-system interaction was also more efficient in producing recasts. System recasts addressed the full range of linguistic errors and specific error types that can be considered relevant for EFL learning. In addition, the fact that system recasts involved errors that were not recast in SCMC indicates that the dialog system may be able to provide feedback on errors that tend to be overlooked by peer interlocutors. While the efficacy of system recasts in terms of noticing and acquisition will need to be tested empirically, the present study revealed that system recasts displayed some characteristics such as the repetition of recasts and target-like forms or a complete focus on learner errors that, from a theoretical point of view, are associated with increased saliency and noticing. Failure to isolate the target-like form in the recast move, however, might in turn mitigate the system recasts' salience. The higher uptake rate of recasts in learner-system interaction compared to SCMC may be considered an initial indication of increased noticing in this interaction type and also points to an advantage in terms of practice and awareness of the recast form.

5.2.3 Lexical Alignment

Lexical alignment in both interaction types occurred from learner to learner (self), from learner to interlocutor, from learner to both as well as from interlocutor to learner. Alignment from interlocutor to interlocutor was excluded from the analysis for lack of interpretative value.

Without correcting for an unequal number of messages between the two interaction types, learner-system interaction led to considerably more overall reuse of lexical material both across all alignment directions as well as with respect to the individual alignment directions. This was likely mainly due to the general efficiency of

learner-system interaction in producing more text in less time on task compared to SCMC, as pointed out above, which at the same time increased the opportunities for lexical alignment.

When corrected for number of messages, the difference was less pronounced, with learner-system interaction leading to 1.5 more overall lexical reuse per message than SCMC. Focusing on the individual directions of alignment, learner-system interaction had higher alignment per message from learner to learner (self) and from learner to both. Learner-system interaction also produced more unprimed (i.e., beyond-threshold) reuse of lexical material by learners and interlocutor. On the other hand, alignment per message from learner to interlocutor and from interlocutor to learner was higher for SCMC than learner-system interaction. Possible reasons for the different degrees of lexical alignment between the two interaction types will be discussed in the following paragraphs considering the qualitative and quantitative findings on each direction of lexical alignment.

Lexical alignment from learner to learner (self) was the proportionally most frequent type of learner alignment in both interaction types. In learner-system interaction, there was a clear prevalence of alignment from learner to learner (self) compared to the other directions of learner alignment, i.e., from learner to interlocutor and from learner to both, whereas in SCMC, alignment from learner to learner (self) and from learner to interlocutor was more balanced. This indicates that learners in the system group seemed to be particularly inclined to reuse their own lexical material but relatively less their interlocutor's compared to learners interacting with a peer through SCMC who aligned to their own and the interlocutor's lexical material to a similar degree.

The dominance of alignment from learner to learner (self) over alignment from learner to interlocutor in learner-system interaction may be explained as follows: As seen in the qualitative findings, a first context that involved alignment from learner to learner (self) in both interaction types were reference shifts, i.e., a learner's quasi-repetition of a previous request while shifting its reference as in, e.g., "and how much cost the room Alpha?" and "and how much cost the room Gamma?". The second context of alignment

from learner to learner involved negotiation of meaning (in learner-system interaction) and learners' restating, specifying, or correcting of previous statements (in SCMC). Reference shifts might have served learners in both groups as an economical way to build utterances by reusing parts of their previous formulations. When occurring in negotiation of meaning and learner restatements, specifications, or corrections, respectively, learners seemed to align to their own previous lexical material to achieve communicative success. Thus, in the case of learner-system interaction, learners might have reused lexical material from their own production in reaction to the interlocutor's indication of non-understanding trying to resolve the communication issue with lexical material that had worked previously with the dialog system. In the SCMC group, learners seemed to reuse previous lexical material preventatively rather than reactively, i.e., to prevent misunderstandings or to prevent information to get lost in the conversation. Since non-understanding and negotiation of meaning were much more frequent in learner-system interaction, alignment from learner to learner (self) in this interaction type was higher accordingly.

As a note, this explanation would indicate that alignment from learner to learner in the context of FL learner-system interaction and SCMC might be, at least partly, a conscious, strategic tool rather than a fully subconscious and implicit automatic process. As mentioned above (see section Alignment in chapter Literature Review), the same has been suggested for L2 and FL interaction in general, as L2 and FL speakers' automatic alignment might be impaired due to a lack of proficiency and they may align more strategically, for instance, to facilitate learning or to enhance understanding (Costa et al., 2008). In SCMC, specifically, it has previously been found that learners use lexical alignment to enhance understandability by their interlocutor (Uzum, 2010), and that conscious, strategic processes do seem to play a role in L2 and FL interaction besides implicit, automatic processes (Michel & O'Rourke, 2019).

Given the much higher rate of negotiation of meaning in learner-system interaction compared to SCMC and thus opportunities for alignment from learner to learner (self) in learner-system interaction, it is surprising that the alignment rate from

learner to learner (self) was only slightly higher in learner-system interaction than in SCMC. This observation needs to be discussed together with the finding that there was a considerably higher amount of unprimed (beyond-threshold) learner reuse of lexical material in learner-system interaction compared to SCMC. In fact, unprimed learner reuse in system interaction was as high as within-threshold learner alignment. In comparison, unprimed learner reuse in SCMC was much less frequent compared to within-threshold alignment. In addition, unprimed learner reuse in learner-system interaction occurred in the same contexts as within-threshold alignment from learner to learner (self), i.e., in the context of negotiation of meaning and reference shifts. Considered together, these observations suggest that the frequency of alignment from learner to learner (self) in learner-system interaction might in fact have been higher than was measured in the present study. A possible methodological reason could be that the five-message threshold in learner-system interaction (vs. the ten-message threshold in SCMC) was set too low. Put differently, the priming effect of lexical material could in fact have lasted longer, i.e., closer to the SCMC ten-message threshold which seemed to measure actual alignment more adequately and was not merely dependent on the visibility on the screen. Similarly, in their study on lexical alignment, Michel and Cappellini (2019) reached the conclusion that the setting of an arbitrary threshold was not conclusive to establish the distance between primes and targets.

As a further difference in the frequencies of alignment between the two interaction types, alignment per message from learner to interlocutor was higher in SCMC than in learner-system interaction, as stated above, although, as just mentioned, the threshold might have been inadequately set in learner-system interaction and the actual number of alignment from learner to interlocutor might have been higher and, counting the instances currently coded as *unprimed*, would possibly have surpassed that of SCMC. Nonetheless, as also stated above, there was a relative infrequency of alignment from learner to interlocutor compared to alignment from learner to learner (self) in learner-system interaction that was not observed in SCMC, indicating that learners in SCMC indeed seem to have been more prone to align to their interlocutor than learners aligning

to the dialog system. Relating this finding back to previous research, there is mixed evidence in the literature on general human-system interaction: While some studies suggest higher alignment from human to dialog system compared to from human to human as human users tend to ascribe lower capabilities to the system, others suggest the opposite as alignment relates to politeness and expression of community membership (H. Branigan & Pearson, 2006). The only study on alignment in EFL learner-system interaction by Sinclair et al. (2019) also found lower alignment from learner to dialog system than from learner to human tutor, as learners felt they received less support from the dialog system than from the human tutor, according to the authors.

These explanations may help to interpret the relatively low frequency of lexical alignment from learner to interlocutor found in the present study: Considering the qualitative finding, lexical alignment from learner to interlocutor in both learner-system interaction and SCMC primarily occurred in adjacency pairs as well as (more rarely) in the context of reference shifts. Adjacency pairs included question-answer pairs and thanking and leave-taking. When involved in question-answer pairs, learners might have aligned to their interlocutor to model their response after the interlocutor's to enhance mutual understanding. There was some evidence for this assumption in the learner interviews, where one learner from each group reported that they reused their interlocutor's terms "projector" and "equipment", respectively, to make themselves understood by the interlocutor. In the case of learner-system alignment, this might have been due to the learner's belief that the system is unable to understand language that it does not produce itself (Costa et al., 2008). In the case of thanking and leave-taking adjacency pairs, learners' alignment to the interlocutor's production could be seen as a sign of increased cooperation with the interlocutor and an effort to match their politeness. Since learners interacting with the dialog system might have seen it less necessary or were less inclined to express cooperation and politeness, the rate of alignment from learner to interlocutor in learner-system interaction was lower overall (H. Branigan & Pearson, 2006). As an additional note, this explanation along with the fact that learners explicitly commented on these instances of lexical alignment in the

interviews again point to a more conscious processing of lexical alignment, as already suggested above.

As a final direction of learner alignment, a small amount of alignment from learner to both could be observed in both interaction types. Since in this alignment direction, learners reused lexical material that had been used by themselves as well as their interlocutor within the threshold of messages, the material involved in alignment from learner to both can be considered particularly prone to alignment. Qualitative findings showed that in learner-system interaction the lexical material involved in alignment from learner to both related to booking confirmations such as dates, while in SCMC it involved exclusively the phrases “in this room” or “in the room”. So, lexical alignment from interlocutor to both generally involved lexical material specifically related to the task, possibly because learners and interlocutors naturally used them frequently when interacting on the task (Dao et al., 2018).

From an EFL learning perspective on the different directions of learner alignment found in learner-system interaction, the repeated use of lexical material in alignment from learner to learner (self) may benefit learners’ development of automatization and fluency in the use of these forms, while alignment from learner to interlocutor may elicit specific lexical items from the learners. Given the presence of both learner alignment types in learner-system interaction indicates that this interaction type may promote both learning aspects, however, the dominance of alignment from learner to learner (self) over alignment from learner to interlocutor suggests a particular focus on automatization and fluency development. Alignment from learner to both can be viewed as playing a role in both, the development of automatization and fluency as well as the elicitation of forms. While in learner-system interaction of the present study, the overall amount of alignment from learner to both was relatively small and limited to a few task-related expressions, it could be suggested that it provides learners with ‘high-intensity’ practice of task-related vocabulary, specifically.

The final direction of lexical alignment considered was from interlocutor to learner. It was the second most frequent alignment direction in both interaction types. While uncorrected for number of messages, alignment from interlocutor to learner was higher in learner-system interaction than SCMC, the alignment per message rate was lower for learner-system interaction compared to SCMC. Again, there was an additional high amount of unprimed interlocutor reuse, which could indicate an inadequate threshold setting, as suggested above and which, if added to the count, would balance out the alignment rates between the two interaction types.

In qualitative terms, alignment from interlocutor to learner in learner-system interaction predominantly occurred in the context of booking request confirmations as well as (rarely) in leave-taking and thanking. The same contexts could be found in some instances of SCMC alignment from interlocutor to learner, although this direction of alignment in SCMC more typically involved repetitions of room requirements. Unprimed interlocutor reuse, i.e., the reuse of n-grams beyond the threshold, in learner-system interaction involved diverse lexical fields from booking date confirmations, room names, leave-taking and thanking, and other task-related vocabulary. In SCMC, the occurrence of unprimed interlocutor reuse was less varied and, as stated above, comparatively less frequent. So, as an initial observation on alignment from interlocutor to learner and unprimed interlocutor use, it can be observed that it generally involved the reuse of lexical material specifically related to the hotel booking task. A similar observation was made by Dao et al. (2018) on structural alignment in L2 task-based learner-peer interaction, who found that the two different tasks used in the study led to the alignment of different syntactical structures related to the topic of the respective task.

As an additional observation on the SCMC interaction type, in the instances where the interlocutor, i.e., the peer in the role of the hotel employee, repeated the room requirements back to the learner (i.e., the client), seemed to represent a kind of sales strategy in that, by repeating the clients' requirements word-by-word, the hotel employee seemed to emphasize that they had exactly what the client needed. This can be seen in

the following extract where the peer aligns to the learner's formulations "a relax situation", "can work in groups", "nice impression", and "device for recording and audios":

- 28 Learner which one of these rooms are the best for 150 people, a relax situation, so that we can work in groups, nice impression and a device for recording and audios. the cost sholud'nt be over 1500 CHF
- 29 Peer the best room is alpha room, 270 people its a relax situation, they can work in groups and a nice impression and device for recording and audios

In learner-system interaction, on the other hand, alignment from interlocutor to learner was coincidental rather than motivated by the interlocutor for any strategic reasons. As explained above (see section Dialog System

), the dialog system's utterances were pre-scripted and did not adapt to the learners' production as the human peer interlocutor could. Firstly, the fact that alignment from interlocutor to learner in learner-system interaction was coincidental rather than motivated or specifically integrated as part of the system design likely led to it occurring less frequently compared to SCMC. Secondly, the observation that alignment from interlocutor to learner in SCMC was used in a very different way, as described above, indicates that human interlocutors as opposed to the dialog system were able to empathize with their customer, identify their thoughts and needs, and flexibly adapt their responses accordingly. The impact that these differences in the alignment from interlocutor to learner in learner-system interaction and SCMC might have on FL development, in terms of the acquisition of forms but also in terms of discourse management, will need to be investigated further.

As a final remark on alignment from interlocutor to learner, the occurrence of lexical alignment from system to learner observed in the present study contrasts with the finding by Sinclair et al. (2019) who observed no alignment from system to learner in their data. The authors explained the absence of this alignment direction with the fact that the used dialog system was not specifically designed to align to the learner. However, the findings from the present study show that lexical alignment from system to learner can in fact occur even in interaction with a system that was not specifically designed for

alignment. The discrepancy between the current study and that by Sinclair et al. (2019) could be due to the fact that the present study used a goal-oriented system that was able to interact on a specific task (i.e., hotel booking) whereas in the study by Sinclair et al. (2019), learners interacted with a chatbot, i.e., a reactive dialog system, on different small-talk topics. Accordingly, learners' and the dialog system's production in the present study involved vocabulary specifically related to the lexical field evoked by the hotel booking task. In the study by Sinclair et al. (2019), on the other hand, learners responded to various small-talk questions by the chatbot on a wide range of topics and, as a reactive system, the chatbot generally relied on generic responses unrelated to the specific content of the learner utterance to be able to handle the diverse user input as in this example: User: "I play volleyball and I go running". Bot: "when do you do that?" (Sinclair et al., 2019, p. 415). Naturally, this type of open-ended dialog on a wide range of topics and using generic chatbot formulations is less likely to produce lexical alignment from interlocutor to learner than task-based interaction on a given topic. In other words, the findings of the present study suggest that if the dialog system along with its language output revolves around a task (here: hotel booking), lexical alignment from interlocutor to learner can occur as learner and dialog system naturally use and reuse vocabulary relating to the topic of the task.

From an EFL learning point of view, the benefits of alignment from interlocutor to learner are to provide learners with positive feedback: By reusing lexical material from the learner's production back to them, the interlocutor implicitly comments on the adequacy of the learner's lexical choice. With respect to the present study, it needs to be pointed out that this function was impacted by the fact that in both interaction types about 40% of all alignment interlocutor to learner involved erroneous language. In the case of system interaction this occurred when the dialog system reused a learner error in booking dates. In SCMC interlocutor alignment may equally reproduce an error from the learner or even introduce an own new error. However, since these errors affected spelling or grammar rather than vocabulary, alignment from interlocutor to learner involving errors may not entirely reverse the effect on lexical development per se.

With respect to the learning potential of alignment from interlocutor to learner, it would also be important to see how the distance from the learner's use of lexical material to the interlocutor's reuse of it impacts the alignment's effect. Specifically, how far from the learner's use of the item may the interlocutor's reuse occur for it to still have a feedback effect? As suggested above, the priming effects in learner alignment might extend beyond the five-message threshold set for the learner-system context and n-grams coded as *unprimed* use might in fact be indicative of alignment. Whether the same applies to alignment from interlocutor to learner and *unprimed* interlocutor alignment beyond the five-message threshold might in fact still be effective would have to be further examined.

In sum, learner-system interaction was more efficient in producing lexical alignment, leading to more overall alignment in less time compared to SCMC, presumably because learner-system interaction led to an overall greater amount of text in less time than SCMC which increased opportunities for lexical alignment. Learner-system interaction involved the same alignment directions as SCMC, with alignment from learner to learner (self) being theoretically linked to the automatization and development of fluent use of lexical items, alignment from learner to interlocutor to the elicitation and adoption of new vocabulary, and alignment from interlocutor to learner to the provision of implicit positive feedback. In both learner-system interaction and SCMC, the overall prevalence of alignment from learner to learner (self) and from interlocutor to learner suggests a possible particular benefit for automatization and development of fluency as well as the provision of positive feedback in both interaction types.

Comparing the amount of alignment per message between the two interaction types, learner-system interaction led to relatively more alignment from learner to learner (self) but less alignment from learner to interlocutor and less alignment from interlocutor to learner than SCMC. So, each interaction type was inherently prone to produce certain directions of alignment. Thus, learners in the system group seemed to have frequently self-aligned to re-establish understanding but might have felt less inclined to align to

their system interlocutor as a display of cooperation and politeness. Alignment from interlocutor to learner was likely lower in system interaction than SCMC as the dialog system's reuse of learner vocabulary was merely coincidental. However, these findings should be considered with caution, as the prime-target threshold setting in the learner-system interaction might have been inadequate and the actual counts could in fact be more balanced between the two interaction types.

As a qualitative finding, the observed lexical alignment generally involved vocabulary specifically related to the task such as for example expressions used in leave-taking and thanking (e.g., "have a nice day") and hotel-related vocabulary (e.g., "book this room", "number of guests"), which suggests that it is specifically task-related vocabulary that is trained and potentially acquired through lexical alignment in learner-system (and SCMC) task-based interaction.

5.2.4 Structural Alignment

Structural alignment involved the same directions as lexical alignment, i.e., from learner to learner (self), from learner to interlocutor, from learner to both, and from interlocutor to learner. Structural alignment from interlocutor to interlocutor was again excluded from analysis due to its lack of interpretative value.

The eight syntactic structures considered for analysis included structures that can be considered highly relevant for EFL learning such as, for instance, modal verb constructions, question forms, or *to*-infinitive. Structural alignment involving these structures may, as has been argued by L2 and FL research (see section Alignment in chapter Literature Review), facilitate the acquisition of linguistic forms in different ways depending on the direction of the alignment: alignment from learner to learner may promote automatization and fluency development when using the structures, alignment from learner to interlocutor may elicit from learners the use of certain structures, and alignment from interlocutor to learner may function as a form of positive feedback by repeating a structure back to learners.

Comparing the structural alignment between the two interaction types, overall structural alignment uncorrected for number of messages was considerably higher in learner-system interaction than in SCMC across all directions and all structures. Since task on time was lower in learner-system interaction compared to SCMC, this means that learner-system interaction compared to SCMC was more time efficient in eliciting more structural alignment in less time.

When corrected for number of messages, the advantage of learner-system interaction in producing higher overall alignment was still there but less pronounced. So, as with lexical alignment (see above), learner-system interaction seems to have been more efficient in producing structural alignment primarily because it produced more text overall, as more text corresponded to higher occurrence of the analyzed grammatical structures and thus opportunities for alignment. This suggestion finds support in the fact that learner-system interaction indeed showed considerably higher occurrences with respect to all structures compared to SCMC, as seen in the overview of total number of occurrences (i.e., all instances of a given structure including alignment and unprimed reuse as well as alignment from interlocutor to interlocutor) of the analyzed structures (see section Structural Alignment in chapter Findings). As the overview furthermore showed, this also applied to occurrences per message for almost all structures, indicating that learner-system interaction also produced text with a denser use of the analyzed structures. This in turn presumably led to the overall slightly higher alignment per message rate in learner-system interaction compared to SCMC.

With respect to the individual alignment directions, there again (see lexical alignment) was a notably higher proportion of unprimed reuse vs. within-threshold alignment in learner-system interaction, which could point to an inadequate setting of the prime-target threshold for learner-system interaction. So, the following differences in alignment frequencies between the two interaction types need to be considered cautiously: While SCMC led to more structural alignment per message from learner to learner (self) and from learner to both, learner-system interaction led to more alignment from learner to interlocutor and slightly more alignment from interlocutor to learner.

Interestingly, this pattern is reverse to that found for lexical alignment (see above), where learner-system interaction led to more alignment per message from learner to learner (self) and from learner to both but less alignment from learner to interlocutor and from interlocutor to learner compared to SCMC. The differential patterns could be linked to the different processing of lexical and structural alignment in learner-system interaction: As explained above, the findings on lexical alignment of the current study suggest that learners seemed to use lexical alignment at least partly as a conscious strategy, for instance, to produce language economically or in response to non-understanding. Also involving conscious strategies, learners appeared to resign from aligning to the dialog system as they considered it unnecessary to express cooperation and politeness in that way (see above). In the case of structural alignment, it could be argued that it involves more subconscious processes and learners automatically aligned to their own structures or those produced by the interlocutor. This argument is based on the observation by Lotze (2016) who observed with respect to native speakers interacting with dialog systems outside the learning context that users tend to reuse grammatical structures unnoticed and without thinking while lexical items are better noticed and might be reused more strategically, particularly in the written context.

The following paragraphs propose what specific processes might have been involved in the structural alignment in learner-system interaction leading to the different proportions of the individual directions of alignment within learner-system interaction and the different rates between the two interaction types. As stated above, learner-system interaction led to more alignment per message from learner to interlocutor but less alignment from learner to learner (self) than SCMC. One explanation for the higher rate of alignment from learner to interlocutor in the system group compared to the SCMC group could be linked to the learners' perceived proficiency of their interlocutor: Research on lexical alignment in SCMC has shown that learners tend to align more if they are interacting with a more proficient interlocutor (Michel & O'Rourke, 2019). In the present study, it can be expected that learners considered the dialog system as highly proficient, which might have driven their alignment to the interlocutor. In the SCMC

group, on the other hand, learners interacted with a peer interlocutor whose proficiency was more or less matched to their own (see section Procedure in chapter Methods), possibly leading to less alignment from learner to interlocutor.

Increased structural alignment from learner to interlocutor in learner-system interaction might also have impacted alignment from learner to learner (self) in the following way: As learners tended to align to the structures modeled by the system interlocutor, they were steered away from structures they had previously used in their own production towards using the system's structures instead, thus decreasing alignment from learner to learner (self). In SCMC, however, learners followed their interlocutor less and tended to reuse structures they themselves had used before.

The suggestion that learners from the system group were steered away from the reuse of their own structures whereas learners from the SCMC group tended to keep reusing their own structures would also explain the finding on the strongly different total occurrences (i.e., all instances of a given structure including alignment and unprimed reuse as well as interlocutor-to-interlocutor alignment) of some structures in the two interaction types: As mentioned above, all eight structures except *have* + noun phrase had higher per message total occurrences in learner-system interaction than in SCMC. This was particularly pronounced for the frequency of questions with subject-auxiliary inversion, which was five times higher in learner-system interaction compared to SCMC, and existential *there*, which never occurred in SCMC.

Interestingly, the structures that stand out (i.e., *have* + noun phrase, question with subject-auxiliary inversion, existential *there*) are all constructions that have an alternative construction that expresses the same meaning. In fact, existential *there* and *have* + noun phrase can be considered mutually alternative constructions to express the same meaning. In other words, existential *there* could be used in place of *have* + noun phrase and vice versa. For example, the sentence *There is enough space in Room Beta* could be alternatively formulated as *Room Beta has enough space*. Similarly, a question may be constructed using subject-auxiliary inversion as in, e.g., *Can you give me the prices?* or as a *wh*-interrogative as in, e.g., *What are the prices?* As the finding on the total occurrences of

the structures suggests, learners in the SCMC group seemed to have consistently preferred one construction, as the alternative while learners in the system group tended to use both alternative structures. The suggestion is that in the system group, the alternative structure was elicited by the dialog system, while such elicitation was absent in the SCMC group. So, learners in the SCMC group exclusively used *have* + noun phrase, driving its total occurrences up to a level above that of learner-system interaction while leading to zero occurrences of existential *there* in SCMC. Similarly, learners from the SCMC group tended to use *wh*-interrogative rather than question with subject-auxiliary inversion, which is reflected in a much lower number of total occurrences of question with subject-auxiliary inversion in SCMC than in learner-system interaction.

As an explanation for the increased alignment from interlocutor to learner in learner-system interaction compared to SCMC it could be suggested that the dialog system's production had a denser use of structures which increased the opportunities for alignment from interlocutor to learner. Support for this argumentation can be found in the higher number of total occurrences of most structures reflecting the overall denser use of the analyzed structures in learner-system interaction. In addition, the considerably higher counts of per message interlocutor reuse (i.e., alignment from interlocutor to learner and unprimed interlocutor reuse) in learner-system interaction compared to SCMC indicate that the structures were used more densely in the dialog system's than the peer interlocutor's production.

In terms of FL learning, comparatively high alignment from learner to interlocutor and from interlocutor to learner in learner-system interaction points to a potential advantage of learner-system interaction with respect to the elicitation of structures through alignment from learner to interlocutor as well as the provision of positive feedback through alignment from interlocutor to learner. On the other hand, the comparatively lower alignment from learner to learner (self) and from learner to both in learner-system interaction points to a possible disadvantage of learner-system interaction compared to SCMC in terms of fluency development.

With respect to the elicitation of structures through alignment from learner to interlocutor, it is important to point out that learner-system interaction might not only have the advantage of leading to more alignment and hence elicitation but also in eliciting structures that were elicited less or not at all through alignment from learner to peer interlocutor SCMC: As shown in the overview of occurrences of the various structures, existential *there* never occurred in the SCMC data and consequently was not aligned to by learners of the SCMC group. However, in learner-system interaction the structure did occur and, as a post-hoc count showed, was also aligned to by the learners in this group. Similarly, question with subject-auxiliary inversion was much more frequent in learner-system interaction than in SCMC, accordingly eliciting the structure more frequently from learners in the system group compared to those in the SCMC group.

To summarize the discussion on structural alignment, learner-system interaction led to a considerable amount of alignment to syntactic structures that are relevant for EFL learning, thus potentially facilitating EFL acquisition of forms by eliciting and providing positive feedback on the structures. While learner-system interaction produced a comparable per message amount of overall structural alignment as SCMC, it was more time efficient in doing so, i.e., it led to more alignment in less time on task. However, as with lexical alignment (see above), these findings should be considered with caution as the five-message threshold in learner-system interaction was potentially too low and within-threshold alignment counts might in fact be higher in learner-system interaction.

Furthermore, while all directions of structural alignment were present in learner-system interaction, it led to more alignment from learner to interlocutor and from interlocutor to learner but less alignment from learner to learner (self) and from learner to both compared to SCMC. In learner-system interaction, the perceived high proficiency of the dialog system might have driven learners to align to their interlocutor while at the same time bringing down the proportion of self-alignment in learner-system interaction. In particular, elicitation through alignment from learner to interlocutor seemed to

involve structures that were aligned to less frequently or not at all in SCMC, as they occurred less frequently or were completely absent in the peer interlocutor's production as opposed to the dialog system's production. With respect to alignment from interlocutor to learner, which was higher in learner-system interaction than SCMC, a generally dense use of the considered structures in the dialog system's production appeared to increase this alignment direction in learner-system interaction. These patterns of structural alignment in learner-system interaction highlight its role in eliciting structures from learners, in particular those structures that would otherwise not be produced by learners, and in providing feedback through alignment from interlocutor to learner.

5.3 Research Question 2

5.3.1 Complexity

Statistical tests revealed a large significant difference between the learners' task performances in the two interaction types with respect to the general syntactic complexity measure *words per AS-unit* and a medium to large significant difference with respect to the lexical diversity measure *MTLD*. The effect on *SYNSTRUT*_t measuring sentence similarity was negligible in size and not significant.

With respect to the *words per AS-unit* measure, learners in the system group produced significantly fewer words per AS-unit compared to SCMC with a large effect size ($d = -0.95$). As described above (see section Complexity Measures in chapter Methods) a high number of words per AS-unit indicates the degree of syntactic complexity in terms of clausal or phrasal elaboration through, e.g., subordination or coordination of clauses, modification, or supplements. So, this finding indicates that learners interacting with the dialog system produced language that was significantly less syntactically complex than those interacting with a peer through SCMC.

Learner responses from the interviews provided additional qualitative findings that may help interpreting this difference. In both groups, learners stated that they focused on

complexity at the sentence-level and indicated that they reduced the complexity either due to their lack of competence or to enhance understanding by their interlocutor. Enhanced clause complexity, on the other hand, was only mentioned once by a learner from the SCMC group stating that they wanted to make longer sentences without specifying any reason. The learners' reported tendency to produce syntactically simple language was reflected in the low mean AS-unit length of $M = 5.48$ words ($SD = 1.29$) in learner-system interaction and $M = 8.42$ ($SD = 4.15$) words in SCMC. In fact, many AS-units in both interaction types consisted of elliptical, minimal sub-clausal units of one to three words (e.g., "Hello", "and room Gamma?", "next time", "fine"), typical of oral interaction as well as SCMC. Syntactic complexity in learner-system interaction might have been particularly reduced since learners from this group had to make an extra effort to make themselves understood by the interlocutor. As mentioned above in the discussion on negotiation of meaning, learners interacting with the dialog system sometimes reduced their messages to single keywords in an attempt to enhance understanding. An additional reason for reduced syntactic elaboration mentioned in interviews by learners from the system group was that they lost patience to build complex sentences.

To sum up the discussion on the different AS-unit lengths in the two groups, learners from both groups tended to focus on reducing rather than increasing the complexity of their sentences. However, while learners in both interaction types produced language with low syntactic complexity and both groups mentioned some of the same reasons for reducing syntactic complexity (e.g., lack of competence, enhance understanding), simple syntax was more pronounced in the system group likely due to the dialog system's lack of understanding and because of the additional reason of learners being annoyed at the dialog system. In terms of its effect on EFL learning, it can be argued that the reduced complexity of the learners' production in the system group might negatively affect their language development in terms of linguistic competence, as learners miss out on producing and practicing more complex syntactic structures.

With respect to the other measure of syntactic complexity, i.e., syntax similarity (measured by SYNSTRUTt), there was not significant difference between the two interaction types. As described above (see section Complexity Measures in chapter Methods), SYNSTRUTt indicates the similarity between the sentences at the clausal, phrasal, and part of speech levels in a text by comparing the syntactic structures of two sentences, combining all sentences across the text. A high SYNSTRUTt value indicates high uniformity and consistency of the syntactic structures in a text, while a low value indicates high syntactic diversity. It might be surprising that the SYNSTRUTt scores did not differ between the two interaction types. Given the high reuse of linguistic material in the context of negotiation of meaning and alignment in learner-system interaction, it could have been expected that this might have also resulted in higher similarity of sentences. One explanation for the absence of an effect might be that SYNSTRUTt analyzes sentence similarity across all sentences of the interaction and not just adjacent ones, where sentence similarity might have been relatively higher in learner-system interaction compared to SCMC. Indeed, post-hoc analysis of the SYNSTRUTa scores (i.e., similarity of adjacent sentences) revealed a large effect for learner-system interaction (system: $M = 0.15$, $SD = 0.06$, SCMC: $M = 0.10$, $SD = 0.05$; $t(37) = 2.55$, $p = .01$, $d = 0.83$), indicating higher similarity of adjacent sentences in learner-system interaction than in SCMC. So, it seems that the negative effect that learner-system interaction did have on syntax diversity of adjacent sentences, likely as a result from high repetition in the context of meaning negotiation, was smoothed out across the entire text.

To sum up the observations on syntactic complexity, learner-system interaction compared to SCMC led to lower AS-unit length, reflecting overall syntactically less elaborate language in learners' task performances of this interaction type. However, learner-system interaction had no effect on syntactic similarity of sentences across the text (measured by SYNSTRUTt), indicating that the syntactic diversity in learners' task performances was overall equal between the two interaction types, even if there was higher syntactic similarity between adjacent sentences in learner-system interaction compared to SCMC.

A second significant medium to large effect was found for lexical diversity measured by MTLTD, with a lower score in the learners' task performances of the system group compared to the SCMC group. As explained above (see section Complexity Measures in chapter Methods) MTLTD calculates the lexical diversity of a text by measuring the mean length of word strings for which a certain type-token ratio is maintained. So, the significantly lower MTLTD measure in system interaction compared to SCMC indicates that learners used less diverse vocabulary when interacting with the dialog system.

In the learner interviews, lexical diversity was only mentioned by one learner from the SCMC group. The learner stated that they reduced the diversity of their vocabulary to enhance understanding and to maintain the speed of the conversation. So, while the learner interviews provide no additional information to account for the difference between the two interaction types, it could be argued that lower lexical diversity in learner-system interaction might have been a result of the learners' frequent self-repetition and rephrasing while reusing lexical material from the original utterances associated with frequent negotiation of meaning in this interaction type.

It needs to be noted that, from an EFL learning perspective, there may be arguments both for and against the benefit of lower lexical diversity: While it might be considered unfavorable for language learning as learners have less opportunity to practice their vocabulary in terms of broadness and more diversity, it might be beneficial in terms of automatization of those items that are used repeatedly.

To sum up the discussion on the linguistic complexity of learners' task performance, learner interacting with the dialog system produced language that was significantly less complex in terms of overall syntactic complexity measured by AS-unit length. While learners from both groups tended to focus on reducing rather than increasing the complexity of their syntax for the same reported reasons, learners from the system group appeared to be more inclined to do so out of frustration over the dialog system's lack of understanding. However, there was no significant negative effect on the

similarity or diversity, respectively, of learner sentences across the text as measured by SYNSTRUTt. It appeared that the frequent repetition of syntactic structures in the context of learner-system negotiation of meaning, albeit measurable, did not impact the syntactic similarity score of sentences across the text. Lexical diversity as measured by MTLT was also significantly lower for learner-system interaction compared to SCMC, possibly as a result of the learners' frequent self-repetition of lexical material in negotiation of meaning. From an EFL learning potential it may be argued that learners interacting with a dialog system might in the long-term miss out on opportunities to produce and practice syntactically and lexically more complex language but might, at the same time, benefit from automatization effects associated with the repeated use of lexical items and syntactic structures.

5.3.2 Accuracy

Small non-significant differences between performances in learner-system interaction and SCMC were found for all accuracy measures except for the request appropriateness measure, which was negligibly and non-significantly lower in learner-system interaction. Thus, learner-system interaction led to a lower rate of errors per AS-unit and a higher rate of error-free AS-units per AS-unit. Conversely, learners interacting with the dialog system produced more inappropriate requests per total requests than those interacting through SCMC. As mentioned above, all effects were small and non-significant, likely due to the limited sample size that was used for this exploratory study. So, the differences need to be interpreted as trends and should be confirmed in further studies with a larger sample size.

The trend towards slightly higher lexico-grammatical and orthographic accuracy in terms of fewer errors and more error-free language in the performance of learners from the system group may be explained based on the learners' questionnaire and interview reports as well as the additional descriptive findings on accuracy: As an initial observation, learners from the system group reported on accuracy more specifically by commenting on different types of errors instead of general accuracy as learners in the

SCMC group did. This could indicate a heightened awareness for accuracy among learners in the system group which might have positively impacted the accuracy of their production.

Common to both groups was the statement that learners focused on accuracy to enhance understanding by their interlocutor. Both groups also mentioned a similar set of reasons for deviating from accurate language use: lack of knowledge, speed, mistyping, and the fact that the interaction took place as part of a low-stakes exercise were named as reasons for producing less accurate language in both groups. In the system group, learners additionally stated that they assumed errors had no impact on understanding and that they lost their patience to produce accurate language because they were annoyed by the dialog system. With respect to SCMC, learners indicated in the interviews that they produced inaccurate language presuming that the interlocutor would let them know if they could not understand them. They also mentioned challenges such as lack of concentration, cognitive overload, and anxiety due to the pressure of interacting with a classmate as negative influences for the accuracy of their production. Furthermore, the 'relaxed' character of the chat modality led them to produce less accurate language. So, the two interaction types shared some characteristics leading to reduced accuracy in the learners' performance and each came with their own additional set of disadvantages, some of which could have additionally negatively impacted accuracy in SCMC, specifically, increased levels of cognitive overload and anxiety, which have been shown to negatively impact FL performance (Zhang, 2019).

A descriptive breakdown of the error types indicated that learner-system interaction showed a higher proportion of grammatical errors, but a lower proportion of spelling errors compared to SCMC. Learners' spelling accuracy in learner-system interaction might have been enhanced compared to SCMC because the dialog system's production was generally more accurate than that of the learner interlocutors, modelling correct spelling to the learner (Nagata et al., 2020). Another reason might be that learners in the SCMC group were more anxious to respond quickly, as indicated in the learner interviews, resulting in more typing errors, which drove the overall spelling error count.

Less anxiety in the system group might also have increased the learners' monitoring as they might have felt more comfortable re-reading their message before sending it off. Since spelling errors tend to be the most obvious error type and are most focused on by learners, they might have been better noticed and corrected more frequently by learners than grammatical and lexical errors. In other words, the system's modeling of accurate language and lower anxiety levels in the system group could have increased the accuracy of the learner production, however, predominantly with respect to spelling errors which were noticed best by learners.

The findings on request appropriateness showed a small non-significant negative effect for the performance of learners in the system group on the number of inappropriate requests per request but only a negligible non-significant negative effect on the mean request appropriateness score. A descriptive overview of the frequencies of the three appropriateness rating levels (i.e., *inappropriate*, *partly appropriate*, *fully appropriate*) also showed that the proportion of inappropriate requests was higher in learner-system interaction than in SCMC. So, in learner-system interaction, learners more frequently produced requests that were too direct or indirect, were insufficiently modified, or were severely erroneous in terms of grammar and/or discourse. Possible reasons for this tendency will be outlined in the following paragraphs.

First, focusing on the degree of request modification, a descriptive overview of the frequency and distribution of modifier types similarly indicates that learners in the system group tended to use less modification overall and, specifically, a smaller proportion of external modifiers, i.e., modifiers outside the head act such as, for example, agreeers (e.g., *That sounds great*) or appreciators (e.g., *Thank you*), compared to internal modifiers, i.e., modifiers inside the head act such as modal preterites (e.g., *We would like to take room Alpha*) or politeness markers (e.g., *Can you please give me the prices?*). This finding is in line with Timpe-Laughlin and Dombi (2020) who investigated learners' request behavior in spoken learner-system interaction and similarly found relatively little external modification. The explanation by Timpe-Laughlin and Dombi (2020) that

learners might have abstained from externally modifying their requests because they were effortful and learners tended to refrain from rapport building and meaning negotiation could be applied to the current study as well: While in human-human interaction, requests are modified to mitigate their face-threatening potential, this might have seemed superfluous to learners as their non-human interlocutor had no face that could potentially be threatened. The fact that some learners were enervated by the dialog system's lack of understanding likely contributed to this tendency. Learners' reluctance to modify their requests particularly affected external modification which is quite 'labor-intensive', here in terms of typing effort. Internal modification, on the other hand, generally involves highly routinized forms such as, *please* or *could you*, which learners were ready to use despite interacting with a non-human agent. In the present study, the learners' use of internal modifiers might have been additionally driven by the dialog system's consistent modeling of these forms. The comparatively high use of modal preterites by learners in the system group, for example, might have been a result from the dialog system's consistent modeling of modal preterite forms. So, overall lower modification of requests in learner-system interaction seems to have been mainly driven by learners' reluctance to use external modification when interacting with the dialog system, as they considered it uneconomical and superfluous when communicating with a non-human interlocutor. Reduced modification in turn might have partly contributed to the trend of more inappropriate learner-produced requests in learner-system interaction compared to SCMC.

With respect to the directness of the requests, the descriptive overview of the distribution of the request strategy types indicates no general tendency for learners in the system group to use a higher proportion of direct requests (i.e., imperatives, direct, indirect, or elliptical questions, *want* statements, assertions) and lower proportion of indirect requests (i.e., references to preparatory condition, strong hints). Instead, learners in both groups generally used the same strategies when realizing requests, with both predominantly using direct questions (e.g., "How much does it cost?") and strong hints (e.g., "I am searching a room for 150 people."). This contrasts with Timpe-Laughlin and

Dombi (2020) who found almost no hints in L2 learner-system interaction. The discrepancy between the two studies could be explained, on the one hand, by transfer from the learners' varying L1s, which has been found to impact learners' request behavior (Timpe-Laughlin & Dombi, 2020), and, on the other hand, by the different learner-perceived capabilities of the respective systems used in the studies in processing learner input. The latter observation indicates that the nature of learners' task performance strongly depends on the specific system used for interaction. However, in the present study, there seemed to be a general tendency in the system group to use more explicit strategies that provide sufficient context, indicated by the slightly higher use of direct and indirect questions, *want* statements, references to preparatory condition, and imperatives, but fewer elliptical questions and strong hints which rely on context-infering by the interlocutor. This reinstates the observation by Timpe-Laughlin and Dombi (2020) that learners tended to rely on explicit request strategies that are not dependent on context or too subtle when interacting with a dialog system. In sum, the trend of more frequent inappropriate requests in learner-system interaction of the present study seemed not to have been a result of learners' preference of direct over indirect request strategies when interacting with the dialog system.

With respect to the impeding impact of linguistic and/or discourse errors (defined here as cohesive issues such as abrupt topic switches or unclear reference, insufficient detail, and issues in the segmenting of sentences) on request appropriateness, the following observations can be made: Given that there was a tendency for higher accuracy in learner-system interaction, more linguistic errors in learners' requests in the system group was unlikely to have led to the lower appropriateness ratings. However, although discourse errors were not specifically analyzed when coding errors, requests in learner-system interaction can be expected to be marked by a relatively high amount of discourse errors: On the one hand, learners in the system group often used 'non-responses' as part of negotiation of meaning sequences to react to the dialog system's indications of non-understanding (see section Negotiation of Meaning in chapter Findings) which naturally also led to requests with abrupt topic switches or unclear reference, i.e., discourse errors.

On the other hand, learners in the system group tended to use more elliptical or even one-word requests, presumably in an attempt to enhance understanding by the dialog system through very simple language (see discussion on complexity). Such requests were rated as inappropriate due to their unclear referencing and/or insufficient level of detail.

So, in sum, the trend to a higher ratio of inappropriate requests per total requests in the performance of learners from the system group was likely caused by learners' tendency to use fewer modifiers, particularly external modifiers, and made more discourse errors in their requests. However, learner requests in learner-system interaction tended not to involve more direct request strategies that would have impeded request appropriateness.

However, despite the higher rate of inappropriate learner requests in learner-system interaction compared to SCMC, there was only a negligible non-significant negative effect for learner-system interaction on the overall appropriateness rating of requests, as stated above. One explanation could be found in the distributions of the *inappropriate*, *partly appropriate*, and *fully appropriate* request ratings in the two interaction types: In learner-system interaction the proportions of fully appropriate and inappropriate requests were higher than in SCMC, while the proportion of partly appropriate requests was lower. Relating the distributions of rating levels in the two interaction types to their overall appropriateness ratings, it could be argued that the relatively high count of fully appropriate requests in system interaction might have balanced out the higher proportion of inappropriate requests, leading to an overall equal appropriateness rating between the two interaction types.

The relatively high count of fully appropriate request in learner-system interaction might be surprising considering the challenges learners met when formulating requests discussed above. One explanation could be that some learners might have been less affected by the dialog system's impeding impact on the appropriateness of requests and, in addition, might have benefited from the system's modeling of consistently appropriate requests. Possible reasons for the observation that the appropriateness of the production of some learners seemed to be positively rather than negatively affected by

dialog system interaction will be discussed in detail in the next section on communicative adequacy.

To summarize the discussion on the findings on the accuracy of the learners' task performances in the two interaction types, learners interacting with the dialog system showed a trend to producing language that was more accurate in terms of lexico-grammar and orthography, i.e., they produced fewer errors per AS-units and more error-free AS-units, than those interacting through SCMC. It needs to be noted that the found effects were small and non-significant and would need to be confirmed in further studies with a larger sample size. Improved lexico-grammatical and orthographic accuracy in learner-system interaction might be partly explained by the learners' self-reported lower levels of anxiety and cognitive load and heightened awareness for accuracy as well as the system's modeling of accurate language, which could all be argued to increase learners' self-monitoring and noticing. These influences might predominantly come into play with the highly visible spelling errors but might be less important with respect to grammar and lexis errors, which were not less frequent in learner-system interaction compared to SCMC.

Conversely, learners from the system group tended to produce more inappropriate requests per total requests than those interacting through SCMC, although the higher rate had no effect on the overall appropriateness rating of requests, presumably because the relatively high rate of fully appropriate requests had a smoothing effect on the overall appropriateness rating in the system group. Again, the effects were small and not statistically significant. Descriptive findings on the request characteristics of the two interaction types suggest that learner requests produced in learner-system interaction were more frequently rated as inappropriate because they tended to contain less modification, particularly through external modifiers, than those in SCMC. Another reason was that learners interacting with the dialog system more often produced requests with unclear reference or involving abrupt topic switches in the context of negotiation of

meaning and elliptical, one-word requests that lacked clear referencing and a sufficient level of detail.

5.3.3 Communicative Adequacy

As outlined above (see section Communicative Adequacy Measure in chapter Methods), the communicative adequacy of learners' performances was measured in terms of task completion and adherence to genre-appropriate style and structure. The findings revealed that learners interacting with the dialog system scored significantly lower on the task completion rating with a medium effect, while the effects on generic style and structure were negligible and non-significant. A descriptive overview of the task completion scores additionally showed that learner-system interaction led to a higher proportion of *unsuccessful* ratings, with almost a fifth of the learners from the system group being unsuccessful in completing the task, and a lower proportion of *successful* ratings compared to SCMC. So, learners interacting with the dialog system were less successful in completing the task than when interacting through SCMC, with a considerable proportion of learners being unable to complete the task altogether, i.e., they did not book a room and/or did not request and process all the necessary information.

The findings from the learner questionnaires and interviews provided some additional insights into possible reasons for the lower task completion rate in the system group. In the questionnaire, more learners from the system group self-reported low levels of task motivation and engagement (e.g., *task was interesting, enjoyed the task, made an effort*). At the same time, fewer learners from the system group agreed with the statement that the task was easy. The same tendencies were found in the interviews where some learners from the system group reported a predominantly negative perception of the task because they found it difficult to communicate and/or thought its usefulness was limited. This suggests that frequent non-understanding in learner-system interaction, also evidenced by the high rate of negotiation of meaning in this interaction type discussed above, might have made the task more difficult to solve for learners and lowered their

success rate in completing the task. Non-understanding and perceived difficulty might also have negatively impacted the learners' motivation and engagement, which could have further contributed to the low success rate.

However, it needs to be pointed out that the majority of learners from the system group were *partly successful* in solving the task (i.e., booked a room after requesting and processing three or four out of five information pieces) and two thirds of the learners interacting with the dialog system gave high ratings on task motivation and engagement aspects as well as the item on the easiness of the task. This indicates, on the one hand, that the task could be solved by interacting with the dialog system and that, on the other hand, some learners appeared to have less difficulty doing so than others.

A possible explanation for the varied success of learners might lie in the strategies they used when selecting a room: The interviews showed that learners from both groups used diverse strategies when choosing a room. Some of the strategies can be expected to help completing the task in either interaction type (e.g., asking about the complete list of requirements one by one, comparing the information on the rooms against the requirements before selecting, excluding rooms that appeared unsuitable), whereas some can be expected to be clearly unhelpful (e.g., selecting a room based on personal prioritization of the requirements, selecting the room before having collected all the required information). Another set of strategies reported by both groups might work in the learner-peer SCMC context but not in learner-system interaction: For instance, stating all requirements at once and ask the interlocutor in the role of hotel employee for a suitable room is a working option for the SCMC context since the human peer interlocutor is able to memorize and process all the information and match it against the three rooms on offer. The human interlocutor will also proactively ask additional information if needed, sharing the responsibility and collaborating with the learner in the role of the client in selecting a suitable room. The dialog system, on the other hand, could only adequately process one piece of information at a time because, in technological terms, correct intent classification was most reliable when the dialog system was presented with a single intent per message (see section Dialog System). Furthermore, the

dialog system was, by design, 'reactive' rather than 'proactive' in that it primarily responded to user requests.

So, learners interacting with the dialog system were required to choose strategies specifically suited for dialog system interaction to successfully communicate and complete the task. Given the above observation that some learners were more successful than others in communicating with the system and completing the task suggests that these suitable strategies were more accessible to them. It would be interesting to investigate in further research what factors (e.g., general digital literacy, previous experience interacting with dialog systems, general communicative competence) might impact learners' use of suitable strategies. As a practical implication, this indicates that learners would generally benefit from knowing which strategies work and which do not when interacting with a dialog system on a task.

From a learning point of view, failure to complete the task might correspond to limited engagement with the complete set of competences targeted by the task. More specifically, learners who were unable to request all required information and/or book a suitable room might have missed out on opportunities to linguistically realize and practice certain requests along with their lexico-grammatical and pragmatic features. In addition, reduced success in completing the task might also have negatively affected the learners' motivation and engagement, as evidenced by the questionnaire and interview findings (see above), which might have a further negative impact on language learning in this group of learners.

With respect to the *generic structure* rating, non-significant negligible differences were found between the two interaction types, with a mean rating of *partly genre-appropriate* structure in both interaction types. A descriptive overview of the ratings also showed that in both interaction types the majority of structures were partly genre-appropriate (i.e., followed the structure from greeting and self-presentation through the decision stage and reservation request to leave-taking with one or two deviations). This

shows that the type of interaction did not impact the degree to which learners adhered to the appropriate structure of the text genre invoked by the task.

As the descriptive overview of the ratings on appropriateness of structure further revealed, the proportion of *inappropriate* generic structure ratings was slightly higher in learner-system interaction than in SCMC, whereas the proportions of *partly genre-appropriate* and *fully genre-appropriate* ratings were comparatively slightly lower in learner-system interaction, possibly because frequent non-understanding by the dialog system led to topic switches as already discussed above (see discussion on negotiation of meaning and task completion), which also affected the overall structure of the interaction.

It might be surprising that learners from the system group did not deviate more from the generic structure given the frequent communication issues in this interaction type. Considering the interview findings, most learners from the system group might have succeeded in adhering to the generic structure despite these difficulties because they actively used strategies supporting an appropriate structure. Thus, learners reported in the interviews that they made sure to stay on topic and resisted interlocutor deviations from the generic structure, for instance by ignoring non-relevant interlocutor messages or by negotiating meaning. These behaviors also became evident in the analysis of learner-system negotiation of meaning where learners appeared to evaluate the relevance of each system indicator of non-understanding (see discussion above on negotiation of meaning). Learners who deviated from the generic structure, on the other hand, reported that they followed no specific strategy when interacting, for instance, because they were unsure what structure would be appropriate, that they decided to provide all information at once, that they followed the interlocutor's (misleading) guidance, that they extended the conversation beyond the required topics, e.g., to test out the limits of the dialog system, or that they needed to 'restart' the conversation due to the interlocutor's non-understanding. Interestingly, largely the same behaviors were also reported by learners in the SCMC group, with exception of the last two behaviors (extended conversation, need for restart), which were absent in the SCMC group. This indicates that learners from both

groups met many of the same challenges with respect to structuring their interaction, which might explain why there was no significant difference in the structure appropriateness between the two interaction types.

The two learner groups also largely overlapped with respect to the *generic style* ratings, in that most performances in both groups showed a mixed style (i.e., used an equal number of formal and informal forms), while the study defined a formal style as *fully appropriate*. It needs to be noted that, as mentioned above (see section Communicative Adequacy Measure in chapter Methods), the rating of the performances' style was based on email and letter samples from the learners' coursebooks. This procedure allowed to rate the learners' performance in line with their classroom instructions and curriculum and was adequate to compare the appropriateness of the performances between the system and SCMC groups. As a limitation to this procedure, the style rating of the learners' performances does not adequately describe how they would be perceived in real-world chat communication. Thus, as chat communication can be argued to require a less formal style than formal business letter communication does, the learners' mixed style in the real world, outside the context of this study, might be considered fully rather than moderately appropriate.

Comparing the rating breakdowns between the two interaction types, in learner-system interaction, mixed style was somewhat more dominant than in SCMC, where a proportionally higher amount of both *informal* and *formal* ratings were observed. The learners' general use of both formal and informal forms in their production might be partly explained by the fact that learners in both groups were uncertain to some degree about what style would be suitable for the task: In the questionnaire, learners from both groups chose mostly, but not consistently, formal characteristics to describe the style that they considered most suitable. In the interviews, some learners from both groups similarly expressed that they thought a mixed style was suitable or that they were at times unsure if more formal or informal style would be suitable. The uncertainty of at least some learners with respect to what style would be appropriate was likely due to the chat

context which learners might associate with an informal setting, as pointed out by some learners in both questionnaire and interviews.

Interestingly, however, learners from the system group seemed to be more certain that a formal style would be appropriate for the task. This became evident in the questionnaire where learners from the system group more consistently chose formal characteristics to describe the suitable style. Moreover, they tended to be more aware of the informal character of their own production as well as of the divergence between the formality of their own and the ideal production: Whereas in the system group, learners selected a clear majority of informal characteristics to describe their own language, learners from the SCMC group selected only slightly more informal than formal characteristics. Similarly, when asked about the match between their own production and that suitable for the task, in the system group, a clear majority of learners considered their own style as not suitable for the task, whereas in the SCMC group, a weak majority of learners thought their style was in fact suitable. Learners from the system group were likely more aware of the informal character of their own production as well as the formal style that the task would have required because of the pronounced formality of the system interlocutor's production. Learners from the system group very clearly characterized the system's production with formal features, whereas learners from the SCMC group selected mostly informal features to describe their interlocutor's style. So, while the dialog system's formal style appeared to have heightened the learners' awareness for the formal style that the task required as well as the discrepancy to their own production, this evidently did not result in a more formal production in this group.

When asked in the questionnaire about this contradiction, most learners from both groups indicated that they lacked the English skills to produce a more formal style and that the chosen style was more effortless to them. Individual learners from both groups additionally indicated that the chosen style seemed more suitable for the chat modality or that they were unsure what style would be suitable. Individual learners from the SCMC groups also mentioned that they chose a more informal style to accommodate the interlocutor (i.e., to be able to write more comprehensibly and faster) or because it

was only a simulation rather than a real dialogue. Two learners from the system group mentioned that they were unmotivated to produce a more formal style due to the interlocutor's non-understanding. These responses, on the one hand, confirm that in both groups the chat context created some degree of uncertainty about the level of formality that would be suitable and indicate that each interaction type might have encouraged a more informal style in its own way. On the other hand, the responses suggest that, even if learners agreed that a formal style would have been suitable, they still produced a more mixed style due to lack of proficiency and/or to avoid effort. From a teaching point of view, these tendencies might be partly counterbalanced by explicitly requiring a formal style in the task and by scaffolding formal forms especially to less proficient learners. Modeling the formal forms in the system production alone seemed to be insufficient scaffolding even if learners recognized the formal features in the production, as discussed above.

As a more general practical implication, the fact that learners shared many of the same uncertainties and challenges regarding the structure and style of the interaction suggests that they need clear instructions and guidelines on how to appropriately structure an interaction and behave stylistically when chatting in the professional context, be it with another human or with a non-human interlocutor. As mentioned above (methods), the learners' curriculum and coursebook informing the present study included no instructions or models on chat interaction. As it can be expected that the use of this form of communication in the professional context will increase in the future, instructing commercial employee apprentices on the production of the genre *chat* would be a useful addition to the commercial employee EFL curriculum.

Summarizing the discussion on the three communicative adequacy measures, task completion was significantly lower in the performance of learners interacting with the dialog system than those of the SCMC group, with some learners being unable to complete the task when interacting with the dialog system. Failure to complete the task was likely caused by the dialog system's lack of understanding and might have been

further aggravated by decreased learner motivation also caused by the system's non-understanding. However, it appeared that those learners who chose suitable strategies when interacting with the dialog system might have been unaffected by these influences and successfully completed the task, which indicates that learners might benefit from a better understanding of how the system works and how it responds to different strategies.

Negligible non-significant differences were found for the two other measures, i.e., generic structure and generic style. With respect to generic structure, most learners' task performances in both interaction types were *partly appropriate*, as learners from both groups seemed to have met many of the same challenges with respect to structuring their interaction and chose similar strategies to respond to them.

Similarly, the majority of learners from both groups chose a mixed style rather than a consistently formal style. This appeared to be partly due to learners in both groups being uncertain about what kind of style would be suitable for the chat context. Other reasons seemed to be lack of proficiency and avoidance of effort in both groups as well as additional interaction-specific reasons such as to accommodate the interlocutor in SMC or lack of motivation due to the interlocutor's non-understanding in learner-system interaction.

The fact that the dialog system's consistent use of formal forms appeared to have heightened the learners' awareness of the required formal style as well as the informal character of their own production but not the learners' actual use of formal style, indicates that uncertainty about the suitable style was not the only decisive factor for the use of informal forms but that other factors, such as lack of proficiency, might have been at least equally important. To support the learners' use of formal rather than informal style in learner-system interaction, the task should more explicitly request formal language and provide scaffolding of formal forms that goes beyond modeling it in the dialog system's production, including in classroom instructions and coursebooks.

6 Conclusion

This final chapter reviews the major findings in light of the study purpose and comments on its implications for EFL research and teaching. The chapter first reviews the study purpose and the major findings with respect to the aspects of learners-system interaction it set out to explore. It then discusses what these findings implicate for the use of dialog systems in language classrooms. Next, the chapter discusses the limitations of the study as well as its contributions for research along with suggestions for further research. The chapter concludes with a reflection on the present and future state of dialog systems for language learning.

6.1 Review of the Study Purpose

The purpose of this study was to investigate task-based interaction between EFL learners and a dialog system to explore its potential for EFL classroom learning and teaching, specifically in the context of the commercial VET program where maximizing opportunities for learner interaction, including with new technology, is central due to limited classroom time and a self-organized learning approach. The study focused on the occurrence and characteristics of interactional features as well as the quality of the learners' task performance to contribute insights on the potential benefits of learner-system interaction for the development of linguistic forms and pragmatic and sociolinguistic competencies, respectively. To highlight the characteristics of the features and performances specific to the learner-system interaction type, they were compared to those occurring in written learner-peer SCMC.

6.2 Key Findings

6.2.1 Interactional Features

The study identified three interactional features in learner-system interaction: negotiation of meaning, recasts, and lexical and structural alignment. The occurrence of

all three features was high and generally higher than that in learner-peer SCMC, despite less time on task in the learner-system interaction context. In terms of the features' nature, they displayed some characteristics that were specific to the interaction type and differed from those found in the SCMC data as well as human-human L2 and FL interaction documented in the literature.

Negotiation of Meaning

The interactional feature negotiation of meaning in learner-system interaction was predominantly initiated by the dialog system rather than the learner. Furthermore, learner-system negotiation of meaning generally had an unspecific indicator, i.e., it did not identify what element of the learner utterance triggered the non-understanding, and it produced indicators that were more uniform and less adapted to the context of the negotiation routine than those produced in SCMC. Moreover, the dialog system may also indicate non-understanding when the learner utterance was in fact target-like. The learners seemed to evaluate the relevance of the system's non-understanding and, if deemed relevant to resolve, responded to the indicator using a range of linguistic modifications to adapt their original utterance. The learners' response strategies were based on their conception of what might have caused the system's lack of understanding. None of the learner-system negotiation routines involved a reaction to response component explicitly indicating that the non-understanding had been resolved, as previously observed in human-human negotiation of meaning. Finally, learner-system negotiation of meaning was often multi-layered, i.e., it involved multiple iterations of attempts to resolve the non-understanding and were more often unsuccessful in resolving the non-understanding compared to the learner-peer negotiation of meaning in SCMC. More inefficient negotiation of meaning in learner-system interaction was caused by the dialog system's limited language processing and production abilities.

Recasts

System recasts of learner errors addressed the full range of error types in terms of linguistic focus, and they also involved errors that were not recast by peer interlocutors in SCMC. Furthermore, the dialog system sometimes recast the same learner error multiple times within a given interaction and repeatedly provided the target-like form of a learner error within the same recast move. Finally, the recast form provided in the recast did not occur in an isolated form but was embedded in longer system messages or sometimes even multiple messages.

Alignment

Lexical alignment in learner-system interaction showed the same general prevalence of alignment from learner to learner (self) and from interlocutor to learner that was observed in SCMC. However, comparing the per-message rates of the individual alignment directions between the two interaction types revealed that learner-system interaction led to more alignment from learner to learner (self) but less alignment from learner to interlocutor and from interlocutor to learner than SCMC. The concentration of alignment from learner to learner (self) in learner-system interaction might be tied to the learners' frequent attempts to re-establish understanding by reusing lexical material. Lower alignment from learner to interlocutor might be indicative of the learners' reluctance to display cooperation and politeness to a non-human interlocutor. Alignment from interlocutor to learner was likely lower in learner-system interaction than SCMC as it occurred merely coincidentally in the dialog system's production. In qualitative terms, the observed lexical alignment involved vocabulary related to the lexical field of the hotel booking task.

Structural alignment, too, involved alignment in all directions in both learner-system interaction and SCMC. Learner-system interaction led to comparatively more structural alignment from learner to interlocutor than SCMC, including alignment to specific structures to which learners in the SCMC group did not align, possibly as learners perceived the dialog system as a highly proficient interlocutor and used its modeled

structures. More alignment from interlocutor to learner in learner-system interaction compared to SCMC might have been driven by the dense use of the analyzed structures in the system's production. However, there was comparatively less alignment from learner to learner (self) in learner-system interaction, possibly as strong alignment from learner to interlocutor diverted learners from reusing own structures to using the dialog system's modeled structures instead.

As becomes evident in the above summary, the patterns of alignment directions in the two interaction types were reversed between lexical and structural alignment: While lexical alignment involved more alignment from learner to learner (self) but less alignment from learner to interlocutor and from interlocutor to learner in learner-system interaction compared to SCMC, structural alignment in learner-system interaction involved comparatively more alignment from learner to interlocutor and from interlocutor to learner but less from learner to learner (self). This could indicate that lexical alignment in FL learner-system interaction is used as a conscious strategy by learners while structural alignment involves a subconscious, automatic process, as lexical items are consciously noticed by learners while grammatical structures are not.

With respect to the findings on both lexical and structural alignment, the methodological limitation concerning the setting of the prime-target threshold in the two interaction types, as discussed below, must be considered.

6.2.2 Learners' Task Performance

The analysis of differences in learners' task performance in terms of complexity, accuracy, and communicative adequacy revealed several differences between the two interaction types.

Complexity

Learners from the system group produced language that was significantly less complex in terms of general syntactic complexity as measured by AS-unit length with a large effect size and in terms of lexical diversity as measured by *MTLD* with a medium to

large effect size. However, the performances of the learners in the two groups differed negligibly and non-significantly in terms of sentence similarity, if considering all sentences of the text, as measured by *SYNSTRUTt*.

Significantly lower overall syntactic complexity (i.e., shorter AS-unit length) in learner-system interaction appeared to have been a result of learners' attempts to make their requests more understandable by the dialog system which often misinterpreted their input. Lack of motivation due to the system's non-understanding might also have contributed to decreased syntactic complexity in this interaction type. Similarly, lower lexical diversity in learner-system interaction compared to SCMC as measured by *MTLD* might also have been result of the learners' frequent self-repetition of lexical material in negotiation of meaning.

However, no difference in the *SYNSTRUTt* scores between the two interaction types suggests that the frequent repetition of learner utterances in the context of learner-system negotiation of meaning did not impact the overall syntactic similarity score of sentences across the text.

Accuracy

Learner-system interaction had a small non-significant positive effect on the accuracy of learners' performances in terms of general lexico-grammatical and orthographic accuracy. Thus, learners interacting with dialog system produced fewer grammar, spelling, and lexis errors per AS-unit and more error-free AS-units per AS-unit. Conversely, learner-system interaction had a small non-significant negative effect on one of the two pragmatic appropriateness measures: Learners from the system group produced more inappropriate requests per total requests. However, the effect on overall request appropriateness was negligible in size and non-significant.

The trend towards slightly improved lexico-grammatical and orthographic accuracy in the learner-system group could have been a result of learners' heightened awareness for accuracy when interacting with the dialog system. Accuracy could furthermore have been supported by the system's modeling of correct forms and possibly

lower levels of anxiety and cognitive load than in SCMC. It could be argued that all these influences raised learners' self-monitoring and noticing in learner-system interaction above that in SCMC, leading to overall more accurate production in the former interaction type. Descriptive findings suggest that this effect applied predominantly to spelling, but not grammar and lexis errors, possibly as spelling errors are most noticeable to learners.

The trend to a higher rate of inappropriate requests in learner-system interaction might have been linked to the learners' reluctance to use modifiers, particularly the more effortful external modifiers, to mitigate the requests' illocutionary force while interacting with a non-human agent. Frequent negotiation of meaning in learner-system interaction also seemed to have impacted request appropriateness, as learners sometimes chose a 'non-response' in reaction to the dialog system's indication of non-understanding, leading to inappropriate requests as they lacked reference to the preceding utterance or involved abrupt topic switches. Furthermore, as learners attempted to enhance understanding by using very simple language, they sometimes produced elliptical, even one-word, requests lacking clear referencing and a sufficient level of detail. Descriptive findings indicate that learner-system interaction led to a relatively higher proportion of ratings at both the low and high end of the request appropriateness rating scale which could indicate that some learners were particularly affected by these influences while others were unaffected and might even have profited from the system's modeling of consistently appropriate requests.

Communicative Adequacy

With respect to the communicative adequacy of learners' performances there was a medium negative effect of learner-system interaction on the *task completion* measure. Some learners in the system group seemed unable to complete the task due to the system's lack of understanding and, likely related, lack of learner motivation and engagement. However, other learners interacting with the dialog system appeared

unaffected by these challenges as they chose adequate strategies to communicate and successfully completed the task as prompted.

Negligible non-significant differences were found on the other two communicative adequacy measures, i.e., *generic structure* and *generic style*. With respect to the generic structure ratings, learners from both groups appeared to face many of the same challenges when structuring their chat interactions, and they also used largely the same strategies to respond to them. Similarly, learners from both groups mixed formal and informal forms, as they appeared uncertain about what style would be suitable for the chat context, lacked the language skills, or wanted to avoid effort. Learners from the system group additionally reported lack of motivation due to the interlocutor's frequent non-understanding. While the dialog system's modeling of consistently formal style seemed to have increased the awareness of learners in this group that a formal style would have been more appropriate, this potentially positive influence on style seemed not strong enough to override the mitigating factors such as lack of proficiency, among others.

6.3 Implications for Teaching

Overall, this study strengthens the case for the potential of dialog systems as virtual interaction partners at least for some aspects of FL learning, while also pointing to some important implications and limitations for their use and efficacy in FL classrooms.

6.3.1 Benefits and Challenges of Learner-System Interactional Features

With respect to the potential to promote the acquisition of linguistic forms through interactional features, learner-system interaction in the present study appeared an efficient way of exposing learners to frequent opportunities to engage in interactional features and benefit from their effects. The findings on the individual features summarized above further indicate their specific benefits and challenges for the development of FL forms.

Negotiation of Meaning

With respect to the feature negotiation of meaning, the fact that negotiation episodes were predominantly initiated by the dialog system highlights the feature's benefit in providing learners with opportunities to produce modified output rather than its role in providing learners with feedback on their production. Moreover, since the dialog system's indication of non-understanding often failed to function as specific and reliable feedback, learners may mainly benefit in terms of automatization and fluency rather than accuracy development, which would require more accurate feedback on their (modified) output. As a further implication, learners receiving feedback from a dialog system may need to rely primarily on self-monitoring when judging the appropriateness of their production instead of the interlocutor's indication. To support learners in reviewing the adequacy of the dialog system's feedback and of their own production, learners may benefit from knowledge on how the system 'understands' their input, i.e., how it assigns or mis-assigns messages to intents, and what kind of discourse strategies might impede or enhance the dialog system's 'understanding' of their input. Such knowledge of the dialog system's functioning might also enhance the learners' success in resolving non-understanding as they were found to generate their responses based on their (mis)conceptions of how the dialog system processes their input. As an additional benefit, improved communication as well as a better understanding of the dialog system's capabilities can be expected to impact positively on learners' motivation.

Recasts

The occurrence and sometimes repeated occurrence of the interactional feature recasts in learner-system interaction showed that dialog systems, even without an explicit feedback function, may provide learners with implicit feedback supporting their accuracy development. The finding that system recasts involved all error types (grammar, lexis, and spelling) suggests that learners' development may benefit in all these areas of linguistic accuracy. An implication of the further finding that system recasts involved errors that learners in SCMC did not address, is that when developing a dialog system for

language learning, the system's responses may be seeded with target forms to promote their recasting and acquisition. Similarly, as the target-like forms in the system recasts of the present study tended to be embedded in linguistic material, which might decrease their salience, measures might be taken when developing a dialog system to enhance the salience of the target-like form. One possibility, for instance, would be to implement error detection in the system which makes learner errors visible (e.g., by underlining the erroneous forms) helping learners to notice the gap between the erroneous form in their own production and the target-like form in the system recast.

Alignment

The general prevalence of lexical alignment from learner to learner (self) and from interlocutor to learner over alignment from learner to interlocutor in learner-system interaction (as well as SCMC) suggests that this feature in learner-system interaction might have a particular benefit for the automatization and development of fluency and the provision of positive feedback. The function in terms of automatization and fluency seems to be pronounced in learner-system interaction as it showed particularly strong lexical alignment from learner to learner (self) compared to the other directions and compared to SCMC. The additional finding that lexical alignment generally involved vocabulary specific to the lexical field of the task, indicates that it seems to be specifically target task-related vocabulary. By extension, this means that teachers may use learner-system interaction on a specific task to evoke alignment and thus promote the development of task-related vocabulary.

Since structural alignment in learner-system interaction (as well as SCMC) involved all alignment directions, it may be an effective means to elicit grammatical structures from learners, promoting the repeated production and practice of forms, and providing positive feedback to learners. The finding that learner-system interaction showed higher alignment from learner to interlocutor and from interlocutor to learner than SCMC, suggests that it may be particularly effective in eliciting structures from

learners, specifically including structures that would not be elicited in learner-peer interaction, and in providing learners with positive feedback on their use of forms.

6.3.2 Benefits and Challenges for Task-based Interaction

The findings on the quality of the learners' task performance relate to the potential of using dialog systems specifically for task-based interaction in EFL classrooms. More precisely, the quality of learners' task performance in terms of complexity, accuracy, and communicative adequacy indicates to what extent learner-system interaction may provide learners with opportunities to engage with the linguistic, sociolinguistic, and pragmatic competences targeted by the task.

The fact that there were differences in some of the complexity and accuracy measures between learner-system interaction and SCMC, implicates that the two interaction types vary in the degrees to which they enable learners to engage with the linguistic competences targeted by the task. First, since learners interacting with the dialog system produced language with significantly lower general syntactic complexity (measured by AS-unit length) and significantly lower lexical diversity (measured by MTLT), learner-system interaction may cause learners to miss out on opportunities to produce and practice syntactically and lexically more complex language. At the same time, however, it could be argued that learners interacting with a dialog system may benefit from automatization effects associated with the repeated use of the same lexical material.

A tendency towards improved accuracy (measured by errors per AS-unit and error-free AS-units per AS-unit) in the performance of learners from the system group compared to the SCMC group suggests that learner-system interaction may assist learners in developing more accurate language in terms of lexico-grammar and orthography. On the other hand, the trend for learners from the system group to produce more inappropriate requests per total requests compared to the SCMC group implicates that learners' development of pragmatic appropriateness might be negatively impacted by dialog system interaction.

The lack of a significant difference in the generic structure and style rating between the two interaction types suggests that learners interacting with the dialog system may be able to adhere to the genre evoked by the task and thus engage with the sociolinguistic and pragmatic competences targeted by the task in comparable ways as those interacting with a peer through SCMC. However, learners' uncertainty with respect to the adequate style and structure in the chat context points to the fact that they would benefit from specific guidance (i.e., classroom and coursebook instructions) on the new, but increasingly common text genre *chat* and *chatbot interaction* in the professional context.

However, a significantly lower task completion rate in learner-system interaction compared to SCMC implies that those learners unable to complete task did not engage with the complete set of competences targeted by the task. As a further finding, some learners from the system group were unable to complete the task due to the system's lack of understanding and decreased motivation, while other learners appeared unaffected by these challenges as they chose adequate strategies to communicate. This finding supports the idea above that an understanding of the system's functioning may be important to increase task completion success rate in this interaction type. The same recommendation is provided by Bibauw et al. (2022) who suggest that the success of dialog systems in language classrooms depends on the preparation of the learners and that teachers should provide learners with adequate modeling of speech acts as pre-task treatment.

6.4 Limitations of the Current Study

The findings presented by this study are subject to several important limitations. Firstly, the scope of the present study was largely descriptive in nature, aiming at providing a detailed account of the characteristic of interactional features and task performance qualities in learner-system interaction. Given the limited previous research on the nature of learner-system interaction, this was a necessary contribution to research on dialog systems for language learning as a base for further studies, including intervention studies. Being descriptive in nature, the study used qualitative analysis of the

data. Consequently, a relatively small sample had to be chosen for analysis. This in turn led to the fact that some of the analyzed characteristics could only be observed a limited number of times in the sample, particularly those relating to interactional features in learner-peer SCMC interaction. As pointed out in the discussion of the concerned codes (see sections Lexical Alignment and Structural Alignment in chapter Discussion), some findings therefore need to be interpreted with caution.

The small sample size also decreased the statistical power of the tests used to detect differences in the learners' task performances between the two interaction types. While the power of the test was high enough to detect medium to large effects between the two groups, it may have been insufficient to detect smaller effect sizes. As mentioned in the discussion on learners' task performance, low test power might have been the reason why smaller differences in the accuracy ratings did not reach statistical significance and must be interpreted as trends such as fewer errors per AS-unit ($d = -0.21$), more error-free AS-units per total AS-units ($d = 0.37$), and more inappropriate requests per total requests ($d = 0.34$) in learner-system interaction. Further studies with larger sample sizes will be needed to confirm true differences between the interaction types with respect to these variables.

Another important limitation of the study is that its findings are limited to the specific learning context of its participants and cannot be generalized to other learner groups with different backgrounds, proficiency levels, or educational settings. This limitation extends to the specific task and dialog system that were used to engage learners in interaction. As described in the chapter Methods, the study used a specific task along with a specially developed dialog system. While the dialog system used in this study, as explained in the chapter Methods, represents the type goal-oriented system with contextualized dialog and shares its ability to allow learners to produce language freely within a given task and semantic context with other systems of the same type, it still differs from other systems in ways that can be expected to impact the nature of the resulting interactions, interactional features, and task performances. For instance, the dialog system developed and used in this study was built using a customized set of NLU

and dialog manager components and was trained on specific data and on a specific set of intents, all of which determined the unique way that the dialog system processes and ‘understands’ user input. Similarly, the dialog system’s responses were based on a set of pre-formulated responses that were specifically created tailored to the study participants’ proficiency level and previous learning. As the architecture and functioning of other dialog systems vary from those of the dialog system developed for this study, it can be expected that they will impact the resulting interactions and findings on interactional features and task performance to some degree. It has been found before, for example, that the specific dialog system used may impact the rate and nature of negotiation of meaning (e.g., Satar & Qinghua, 2020) or that learners model their production after the dialog system’s (Nagata et al., 2020). As will be discussed in more detail below, when the present thesis was completed, a new generation of dialog systems relying on large language models had been launched whose capabilities in ‘understanding’ and producing language are radically improved from those of previous dialog systems including the one developed for this study. It can be expected that there will be considerable differences between learners’ interaction with the dialog system developed for the present study and a dialog system of the new generation, most prominently in terms of the rate and nature of negotiation of meaning.

As a further limitation, it needs to be pointed out that the learner-system interactions in the present study were clearly a simulation of human-human EFL learner interactions. This became apparent in several aspects of the learner-system interactions: Most prominently, the dialog system was a virtual interlocutor rather than a living person that merely simulated human behavior such as human cooperation and human language. Thus, the dialog system’s language production, being generated based on a set of pre-defined templates (see section Dialog System

in chapter Methods), was only an approximation of human language production. Most notably, it was much less varied and adapted to the context than that of humans, which became visible, for example, in the dialog system’s negotiation of meaning. The dialog system’s output was also closer to standard written English than would be

expected from a human interlocutor in the chat context. As already mentioned as a limitation above (see section Dialog System

in chapter Methods), modeling the dialog system's responses after email and letter samples in the learners' coursebook ensured that the task was relevant to the participants through using language that was suitable to the learners' language level and provided learners with model input in line with the curriculum's target language norms. The dialog system's production also included the use of lengthy messages as well as sending multiple messages in the same turn, which learners were prevented from doing when interacting with the dialog system. These traits in fact seem to be characteristic of dialog system production in general, as the same observations have been made in native speaker human-system interaction with a range of different dialog systems (Lotze, 2016).

Furthermore, as detailed in the findings, the negotiation of meaning observed in the learner-system interactions of the present study differed from those observed in human-human learner interaction in that they were more uniform, less dynamically adapted to the context (see above), and thus, less efficient. While recasts and alignment in learner-system interaction closely resembled those observed in learner-peer SCMC in linguistic terms, they were entirely differently motivated in the two interaction types: Whereas in human-human interaction, interlocutors intentionally provide recasts as a form of corrective feedback and their alignment involves social-psychological processes, they occurred merely coincidentally in the dialog system's production.

The fact that the learner-system interactions were thus clearly simulations of real human-human learner interaction raises two important questions which will be addressed in the first suggestion for further research below: First, can learner-system interaction, being a simulation of human-human interaction, facilitate language learning as human-human learner interaction does? More specifically, do the specific characteristics of learner-system interactional features impact – positively or negatively – their efficacy in promoting development of linguistic forms? And does task performance in learner-system interaction predict learners' development of communicative competence as in human-human learner interaction?

Second, are the communicative competences that learners develop in learner-system interaction transferable to real-life human-human interaction? As mentioned above, learner-system interaction takes place in a particular 'social' context, which can be expected to impact learners' sociolinguistic performance (e.g., use of social variants, politeness, humor) and discourse performance (e.g., turn taking behavior, discourse structuring, coherence and cohesion). While, as mentioned in the chapter Literature Review (see section Summary and Rationale for Choice of Theoretical Framework), the cognitive-interactionist approach used in the present study excluded a discourse and sociolinguistic analysis of the interactions which would allow to systematically analyze these elements, the study did produce some findings that indicate an influence on learners' sociolinguistic and discourse performance, e.g., trend towards more inappropriate requests, different turn-taking behavior, absence of humorous and creative language use. So, learners might develop aspects of communicative competence that are specific to the interaction with a dialog system but might not be transferrable to human-human interaction.

As a final limitation, the dialog system and chat application used in this study to generate learner-system and learner-peer SCMC interactions, respectively, differed somewhat in terms of the visual appearance and usability of its interfaces. Specifically, the two application interfaces used different fonts and font sizes and showed a different number of messages visible on the screen at once. These differences might have impacted aspects such as the learners' self-monitoring and noticing, (spelling) accuracy, and priming effects, and thus the findings on interactional features and task performance with regard to the two interaction types.

6.5 Contributions and Suggestions for Further Research

Notwithstanding these limitations, the study offers valuable insights into task-based learner-system interaction and adds to the growing body of research on dialog systems for language learning.

6.5.1 Contributions

With regard to research on interactional features, the study has provided additional evidence on the presence of the interactional features negotiation of meaning and lexical and structural alignment in learner-system interaction and was the first study to reveal the occurrence of recasts in this interaction type. Specifically, it was the first study to identify these interactional features in task-based dialog-based learner-system interaction, making an important contribution to research on the use of dialog systems in communicative L2 and FL classrooms.

The study confirmed the findings from previous studies that negotiation of meaning and alignment in learner-system interaction show unique characteristics when compared to human-human interaction, and demonstrated for the first time that the same is true for recasts. By providing a detailed description of the specific characteristics of interactional features as they occur in learner-system interaction, the study may serve as a basis for further investigation of the role of interactional features in learner-system interaction for FL development.

With respect to the aspect of learners' task performance, the study added to the very limited previous findings on how the learner-system interaction type influences learners' task performance in terms of complexity and accuracy compared to human-human interaction. Moreover, it was the first study to include communicative adequacy when analyzing learners' task performance in system interaction, thus contributing to the research on the potential of learner-system interaction for the development of sociolinguistic and discourse competence, in particular.

6.5.2 Suggestions for Further Research

An important goal for further research of learner-system interaction for FL learning should be to investigate the effect a learner-system has on the development of learners' communicative competence in the FL. As stated above, learner-system interaction is a simulation of human-human interaction and differs from it in multiple ways. For instance, as found in the present study, the interactional features observed in

learner-system interaction are different from those in human-human interaction. By extension, this means that the findings from human-human learner interaction on the effect of interactional features on the acquisition of FL forms cannot simply be transferred to the learner-system interaction context. Instead, FL research must specifically investigate the effect of the interactional features as they occur in learner-system interaction on the development of FL forms. Similarly, it needs to be investigated how task performance in learner-system interaction predicts learners' development of FL competence. More specifically, do high levels of task performance measures in learner-system interaction predict gains in learners' FL communicative competence?

With regard to both questions, it is important to focus on competence gains in the context of human-human communication rather than communication with dialog systems. After all, FL learning is ultimately interested in the question of how language learners best learn to communicate with other humans in a FL. In other words, research on dialog systems for language learning should focus on the question whether there is transfer of FL forms and competences developed in human-system interaction to the human-human communication context. With regard to learner-system interaction, an interesting question is whether there is transfer specifically of sociolinguistic and discourse competences from the learner-system interaction 'social' context to that of human-human communication.

Also needed to better understand the role of learner-system interaction for language learning are longitudinal studies on learners' long-term and repeated use of dialog systems. This suggestion is based, among others, on the present study's finding that learners' negotiation of meaning seemed to be impacted by their understanding and conceptions of the dialog system's functioning. Appropriate understanding of the dialog system's functioning and choice of adequate strategies when interacting with the dialog system appeared to also have led to more success at completing the task. These observations raise the question whether interactional features such as negotiation of

meaning and learners' task performance change as learners become more familiar with the dialog system's functioning over time.

It will also be important to establish what other factors of learner-system interaction influence the number, nature, and impact of interactional features and the learners' task performance. The present study indicated that additional factors, such as learners' motivation and engagement, as well as their understanding of the dialog system's functioning (see above), may have played a role in the degree and ways in which learners and dialog system engaged in interactional features. Other studies have shown an impact for learners' level of language proficiency, type of dialog system and task on the frequency, characteristics, or effects of interactional features such as feedback (Wilske, 2015) or negotiation of meaning (N.-Y. Kim, 2017b; Satar & Qinghua, 2020). In the present study, learners also stated that a lack of motivation negatively impacted their production in terms of complexity, accuracy, and communicative adequacy. A better understanding of whether and how factors such as, e.g., type of system, task, and learner factors influence the occurrence and effect of interactional features and the quality of learners' task performance will help to optimize the setting, task, and dialog system itself, to enhance the learning-facilitating effects of interactional features and performance in learner-system interaction.

6.6 Future of Dialog Systems for Language Learning: Potential and Risks

When the present study was conducted and documented from 2018 to mid-2023, the use of dialog systems for interaction in FL classroom learning and teaching was limited. This was presumably due to a lack of availability of suitable and accessible dialog systems for teachers and learners as well as the systems' reduced capability of imitating human interaction, particularly task-based interaction. In their respective recent reviews of dialog systems for language learning, Bibauw et al. (2022) and Fryer et al. (2020) counted only a few dialog systems specifically designed for FL learning currently accessible to language teachers: *Mondly* chatbot (ATi Studios, 2016) and virtual-reality-

embedded *ImmerseMe* (ImmerseMe, 2021) which enable interaction in various communicative contexts available for several target languages such as, e.g., at the restaurant, buying a ticket at the museum, speaking to a patient as a nurse. Bibauw et al. (2022) and Fryer et al. (2020) mention additional dialog systems allowing for interaction without focusing on a specific task, i.e., *Andy* (ZTO Labs, 2021) and *Tutor Mike* (Chang, 2023) available exclusively for English, as well as “general-purpose chatbots” (Bibauw, François, & Desmet, 2022, p. 131) such as *Cleverbot* (Carpenter, n.d.) and personal assistants such as *Google Assistant* or *Alexa* by Amazon which may be used in language learning classrooms for open-ended conversations without a specific goal rather than task-based interaction. As becomes evident from this overview, dialog systems to teachers and learners for language learning were still very scarcely available at the time of the study. Moreover, the dialog systems that are available are limited in terms of their thematic scope as well their language understanding and dialog management, restricting their suitability for classroom application.

However, at the time of completing the documentation of this study in 2023, there had been some important advancements in the availability and ability of dialog systems within and without the language learning context. In November 2022, the general-purpose dialog system *ChatGPT* was released by the company OpenAI and has since attracted wide attention including from (language) teachers and learners thanks to its advanced NLU capabilities as well as its ability to generate novel text in multiple languages. *ChatGPT* and similar dialog systems launched by other companies such as, for example, Microsoft’s *Bing* chatbot or Google’s *Bard* are part of a new generation of dialog systems relying on large language models (Yan, 2023), i.e., a type of algorithm that uses deep learning techniques and very large corpus of text data with the goal to process, understand, and generate text that closely imitates human language and include a wide range of text types, topics, and styles such as, e.g., poems, essays, conversations, formal letters and so on. While not explicitly developed for language learning, *ChatGPT* and similar dialog systems may be used for language learners and teachers in various ways such as to explain unknown terms and translate, to model various genres of text, to adjust

the complexity or style of a text, or to provide feedback and corrections for learner texts (Kohnke et al., 2023). It may also act as a virtual conversational partner on different topics and thus be used to engage learners in interaction, however, its suitability to be used in interaction on specific tasks is limited. However, the language learning company Duolingo in collaboration with OpenAI has used the same language model GPT-4 as ChatGPT to launch a new app feature, *Roleplay*, which enables learners to engage in interaction with a virtual conversation partner in a more task-based context, e.g., shop for furniture, order a coffee. It can be expected that other developers of language learning applications will follow the example of Duolingo and launch 'new generation' task-based dialog systems for language learning relying on large language models. Two examples for an endeavor in this direction is the current (2023 – 2024) development of a voice-based chatbot for EFL learners relying on the ChatGPT language model by PH Zürich (PH Zürich, n.d.) or the language learning app *Quazel* (Quazel, 2023) launched in 2023 that offers FL learner-system interaction at its core.

These recent developments highlight what Bibauw et al. (2019) predicted for the future of dialog systems in language learning, namely that as the technologies, approaches, and tools for the development of dialog systems for general use advance, the production of dialog systems specifically for language learning becomes more efficient, thus driving their application and use in language learning classrooms. At the same time, the recently launched dialog systems help to understand and foresee their potential as well as their risks when used in language learning classrooms. These will be outlined in the following paragraphs with a focus on their use for FL learning through interaction.

On the one hand, as the availability and suitability of dialog systems for the language learning context improves, they may more readily be used by teachers as a tool to engage learners in interaction, benefitting learners in their FL development in the ways described previously. With respect to the dialog systems' improved language understanding and dialog management, these can be expected to positively impact teacher and learner acceptance and motivation. Besides the motivational impact, improved system capability also means fewer instances of non-understanding and 'false

negative' feedback. Based on the findings of the present study, more capable dialog systems and higher learner motivation could lead to more effective interactional features, particularly negotiation of meaning and feedback, as well as improved learners' task performance in terms of appropriate pragmatic behavior, complexity, and task completion, thus enhancing the overall learning potential of learner-system interaction. To add to the benefits for FL learning, future FL dialog systems might integrate additional functions, such as adapting to the learners' proficiency level, providing explicit feedback on the learner's language use and performance, or allowing teachers to review the learners' performance.

On the other hand, various risks must be considered with respect to the more common use of dialog systems in language classrooms. As will become evidence in the overview below, some of these risks reflect the general debate around the use of artificial intelligence application in fields outside of education. A first risk concerns the collection and storage of learner data and potential misuse of it. The ethical and legal handling of user data is an important issue for any digital technology that is used in education but seems of particular concern with respect to AI-based dialog systems as learners' text input has the potential of revealing highly sensitive information, including not only students' personal data that they may reveal while interacting, but also emotional states algorithmically extracted from the text, for instance for the purpose of assessing learning motivation (Jiang, 2022). To protect learner data and privacy, teachers are required to understand the dialog systems' data management and choose an adequate application. However, the complexity of the issue and the magnitude of the risk regarding the misuse of sensitive student data require that the responsibility for the safe use of dialog systems and other learning technology in classrooms be shared at all levels of the education system, i.e., teachers, the school, as well as other entities of the education system who should provide teachers with adequate training and resources. Above all, the developers and providers of dialog systems must ensure and be monitored that their applications comply with data protection guidelines.

From another perspective on the issue of privacy, learners knowing that their interaction is not taking place in the comparatively private environment of a classroom as FTF interaction is, but is recorded and processed, might give them a sense of being monitored and influence their behavior, even if the data is anonymized. While student-monitoring at first glance might sound favorable from a teacher's point of view as it may improve the students' adherence to the task, some negative effects could for example be that learners are less creative, take less risks in trying out new formulations or become non-compliant and lose inherent motivation.

Another issue are biases inherent in dialog systems relying on machine learning. This issue is part of the general debate on artificial intelligence applications, as existing applications have been observed to contain biases caused by biases in the training data, algorithms, and user interaction, potentially leading to real-life discrimination (Mehrabi et al., 2021). With respect to dialog systems for language learning, linguistic and other cultural biases, for example, might cause the dialog system to be biased towards a specific dialect or cultural concepts. Such biases might, for example, restrict the learners' exposure to different varieties of the FL or perpetuate discriminating stereotypes (Kohnke et al., 2023).

As a further issue, also discussed in the general field of artificial intelligence, the ethical consequences of using a virtual conversation partner for human interaction must be considered. Research on artificial intelligence has observed that humans tend to anthropomorphize technological artifacts, including dialog systems (Salles et al., 2020). As dialog systems become more advanced in understanding language and managing dialog, i.e., more 'human-like' in their interaction, the users' illusion of interacting with another human might even grow for users. Possible consequences suggested by research include, for example, that users might be more vulnerable to be manipulated, that they might be more likely to disclose sensitive information with the risks described above, or that their well-being and socialization might suffer from being emotionally invested in a relationship that is meaningless and whose mutuality is only an illusion (Salles et al., 2020). These implications appear especially severe with respect to young learners who are

still developing psychologically. The potential risks of such behavior must be further investigated by FL research on dialog systems and teachers should be aware of them.

Finally, introducing dialog systems to classrooms to substitute human interaction to some degree, could affect the autonomy of teachers and learners in terms of their ability to take control of their teaching (J. Huang, 2005) or their own learning process (Holec, 1981), respectively. The role of teachers transforms from that of a creator and mediator to a facilitator as it is reduced to merely selecting a dialog system whose scope in terms of task, topic, and language use are otherwise pre-defined and perhaps reviewing the learner data, whereas in the context of human-human learner interaction, it consisted of creating tasks, assigning pairs, participating, and giving feedback in interaction. On the one hand, this can be seen as a welcome relief for teachers who are teaching an increasing number of students in their classrooms, as the number of school children grows and that of teachers is reduced (Bundesamt für Statistik, 2022). On the other hand, constraining teacher autonomy might negatively impact learning outcomes, motivation, as well as teachers' professional satisfaction (Lamb, 2008). Similarly, learners may gain autonomy in their language learning by becoming independent from the availability of a teacher, peer, or native speaker to interact in the FL. However, interacting with a dialog system rather than a human partner reduces their control of interaction at the same time, not only in terms of choosing a conversation partner and topic – a lack of control learners commonly experience in teacher-led classrooms – but also their language use (e.g., turn taking, discourse structuring, style, dialect, use of humor) as a result of a dialog system's linguistic bias described above. This not only limits the learners' practice of certain aspects of language use but might also give them a sense of lack of autonomy with potential long-lasting effects on the learners' development of their own language learning competence as well as their attitude and motivation for language learning (Macaro, 2008).

The following thought concludes the current discussion on the future of dialog systems for language learning and this thesis: When asked about the reasons for learning

languages, common responses by learners are to be able to communicate and connect with people and cultures, to become eligible for certain education programs or jobs and be able to contribute better in one's profession, or to gain access to other cultures and develop personal knowledge of the world (Poarch, 2023). In Swiss elementary schools, foreign languages are taught with the goal of enabling students to access and mediate knowledge and learning, to express themselves and communicate with others, and to become active members of society (Bildungs- und Kulturdirektion des Kantons Bern, 2016). So, our purpose for learning and teaching foreign languages is fundamentally humanistic: to connect with other humans and contribute to the development of society and the wellbeing of its members.

As dialog systems become more available and more capable to process and generate language, they can be expected to be able to engage FL learners in interaction and its benefits in similar ways as a human interlocutor would, with the added benefits of allowing learners to access interaction more independently and adapting to the learners' needs. If the potential of dialog systems now and in the future is to be used towards rather than against the humanistic goals of language learning, the control over their development and their role in learning and education need to lie in the hands of humans with the same cause at heart.

References

- Abrams, Z. I. (2008). Sociopragmatic features of learner-to-learner computer-mediated communication. *CALICO Journal*, 26(1), 1–27. <https://doi.org/10.1558/cj.v26i1.1-27>
- Adams, R., Nik Mohd Alwi, N. A., & Newton, J. (2015). Task complexity effects on the complexity and accuracy of writing via text chat. *Journal of Second Language Writing*, 29, 64–81. <https://doi.org/10.1016/j.jslw.2015.06.002>
- Alemi, M., & Haeri, N. S. (2020). Robot-assisted instruction of L2 pragmatics: Effects on young EFL learners' speech act performance. *Language Learning & Technology*, 24(2), 86–103. <https://doi.org/10.10125/44727>
- Alemi, M., Meghdari, A., & Ghazisaedy, M. (2015). The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics*, 7. <https://doi.org/10.1007/s12369-015-0286-y>
- Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, 28(4), 543–574. <https://doi.org/10.1017/S0272263106060268>
- Anthony, L. (2019). *Laurence Anthony's AntConc* (3.5.8) [Computer software]. Waseda University. <https://www.laurenceanthony.net/software>
- ATi Studios. (2016). *Mondly* [Computer software]. ATi Studios. <https://play.google.com/store/apps/details?id=com.atistudios.mondly.languages&hl=en>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Belz, J. A., & Kinginger, C. (2002). The cross-linguistic development of address form use in telecollaborative language learning: Two case studies. *Canadian Modern Language Review*, 59(2), 189–214. <https://doi.org/10.3138/cmlr.59.2.189>
- Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based

- CALL. *Computer Assisted Language Learning*, 32(8), 827–877.
<https://doi.org/10.1080/09588221.2018.1535508>
- Bibauw, S., François, T., & Desmet, P. (2022). Dialogue systems for language learning. In N. Ziegler & M. González-Lloret (Eds.), *The Routledge handbook of second language acquisition and technology* (1st ed., pp. 121–135). Routledge.
<https://doi.org/10.4324/978135117586-12>
- Bibauw, S., François, T., Van den Noortgate, W., & Desmet, P. (2022). Dialogue systems for language learning: A meta-analysis. *Language Learning & Technology*, 26(1), 1–24. <https://doi.org/10125/73488>
- Bildungs- und Kulturdirektion des Kantons Bern. (2016). *Lehrplan 21 für die Volksschule des Kantons Bern*. Bildungs- und Kulturdirektion des Kantons Bern.
<https://be.lehrplan.ch/>
- Bitchener, J. (2004). The relationship between the negotiation of meaning and language learning: A longitudinal study. *Language Awareness*, 13(2), 81–95.
<https://doi.org/10.1080/09658410408667088>
- Blake, C. (2009). Potential of text-based internet chats for improving oral fluency in a second language. *The Modern Language Journal*, 93(2), 227–240.
<https://doi.org/10.1111/j.1540-4781.2009.00858.x>
- Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology*, 4(1), 120–136.
<https://doi.org/10125/25089>
- Blake, R. (2017). Technologies for teaching and learning L2 speaking. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 107–117). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781118914069.ch8>
- Blum-Kulka, S., House, J., & Kasper, G. (1989). *Cross-cultural pragmatics: Requests and apologies*. Ablex Publishing Corporation.
- Bowles, M. A., & Adams, R. J. (2015). An interactionist approach to learner–learner interaction in second and foreign language classrooms. In N. Markee (Ed.), *The*

- handbook of classroom discourse and interaction* (pp. 198–212). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118531242.ch12>
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13–B25. [https://doi.org/10.1016/S0010-0277\(99\)00081-5](https://doi.org/10.1016/S0010-0277(99)00081-5)
- Branigan, H., & Pearson, J. (2006). Alignment in human-computer interaction. In K. Fischer (Ed.), *How people talk to computers, robots, and other artificial communication partners. Proceedings of the workshop on how people talk to computers, robots, and other artificial communication partners* (pp. 140–156). Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition Universität Bremen/Universität Freiburg. <https://nats-www.informatik.uni-hamburg.de/pub/User/KerstinFischer/hriproc.pdf>
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687–699. <https://doi.org/10.1177/001316448104100307>
- Brown, D. (2014). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research*, 20(4), 436–458. <https://doi.org/10.1177/1362168814563200>
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). John Benjamins Publishing Company. <https://doi.org/10.1075/llt.32.02bul>
- Bulté, B., & Roothoof, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. *System*, 91. <https://doi.org/10.1016/j.system.2020.102246>
- Bundesamt für Statistik, B. für. (2022). *Szenarien für die obligatorische Schule—Lehrkräfte*. Bundesamt für Statistik. <https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/szenarien-bildungssystem/obligatorische-schule-lehrkraefte.html>

- Çakmak, F. (2022). Chatbot-human interaction and its effects on EFL students' L2 speaking performance and anxiety. *Novitas-ROYAL (Research on Youth and Language)*, 16(2), 113–131. <https://novitasroyal.org/volume-16-issue-2-october-2022/?wpdmc=volume-16-issue-2>
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 1–47). Routledge. <https://doi.org/10.4324/9781315836027>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/1.1.1>
- Carpenter, R. (n.d.). *Cleverbot* [Computer software]. <http://www.cleverbot.com/>
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5–35. <https://doi.org/10.5070/l462005216>
- Chang, R. (2023). *Tutor Mike* [Computer software]. Rong-chang ESL, Inc. https://www.rong-chang.com/tutor_mike.htm
- Chen, J. C. (2018). The interplay of tasks, strategies and negotiations in Second Life. *Computer Assisted Language Learning*, 31(8), 960–986. <https://doi.org/10.1080/09588221.2018.1466810>
- Chun, D. M. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System*, 22(1), 17–31. [https://doi.org/10.1016/0346-251X\(94\)90037-X](https://doi.org/10.1016/0346-251X(94)90037-X)
- Chun, D. M. (2008). Computer-mediated discourse in instructed environments. In S. S. Magnan (Ed.), *Mediating discourse online* (pp. 15–45). John Benjamins Publishing Company. <https://doi.org/10.1075/aals.3.04chu>
- Collentine, K. (2019). Task completion and its effects on linguistic complexity/accuracy in a 3D world. *Revista de Linguas Modernas*, 31, 165–183. <https://doi.org/10.15517/rlm.voi31.40875>

- Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes*, 23(4), 528–556.
<https://doi.org/10.1080/01690960801920545>
- Cote, S., & Gaffney, C. (2021). The effect of synchronous computer-mediated communication on beginner L2 learners' foreign language anxiety and participation. *Language Learning Journal*, 49(1), 105–116.
<https://doi.org/10.1080/09571736.2018.1484935>
- Council of Europe. (2001a). *Common European framework of Reference for Languages: Learning, teaching, assessment*. Press Syndicate of the University of Cambridge.
- Council of Europe (Ed.). (2001b). *Common European Framework of Reference for Languages: Learning, teaching, assessment* (10th print). Cambridge Univ. Press.
- Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research*. SAGE.
- Crossley, S., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, 59(2), 307–334.
<https://doi.org/10.1111/j.1467-9922.2009.00508.x>
- Cunningham, D. J. (2016a). Request modification in synchronous computer-mediated communication: The role of focused instruction. *The Modern Language Journal*, 100(2), 484–507. <https://doi.org/10.1111/modl.12332>
- Cunningham, D. J. (2016b). Second language pragmatic appropriateness in telecollaboration: The influence of discourse management and grammaticality. *System*, 64, 46–57. <https://doi.org/10.1016/j.system.2016.12.006>
- Dao, P., Trofimovich, P., & Kennedy, S. (2018). Structural alignment in L2 task-based interaction. *ITL – International Journal of Applied Linguistics*, 169(2), 293–320.
<https://doi.org/10.1075/itl.17021.dao>
- De Jong, N. H. de, Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34(1), 5–34.
<https://doi.org/10.1017/S0272263111000489>

- De la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition*, 24(1), 81–112.
<https://doi.org/10.1017/S0272263102001043>
- DeKeyser, R. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19(2), 195–221.
<https://doi.org/10.1017/S0272263197002040>
- DeKeyser, R., Criado, R., & Chapelle, C. A. (2013). Automatization, skill acquisition, and practice in second language acquisition. In *The Encyclopedia of Applied Linguistics*. Blackwell Publishing Ltd. <https://doi.org/10.1002/9781405198431.wbeal0067>
- Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J. A., & Braasch, J. (2021). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning*, 0(0), 1–29. <https://doi.org/10.1080/09588221.2021.1879162>
- Doughty, C., & Pica, T. (1986). “Information gap” tasks: Do they facilitate second language acquisition? *TESOL Quarterly*, 20(2), 305–325. <https://doi.org/10.2307/3586546>
- Egi, T. (2007). Interpreting recasts as linguistic evidence: The roles of linguistic target, length, and degree of change. *Studies in Second Language Acquisition*, 29(4), 511–537. <https://doi.org/10.1017/s0272263107070416>
- Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: Learner responses as language awareness. *The Modern Language Journal*, 94(1), 1–21.
<https://doi.org/10.1111/j.1540-4781.2009.00980.x>
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(3).
<https://doi.org/10.1111/exsy.12667>
- Ellis, R. (1991). The interaction hypothesis: A critical evaluation. In E. Sadtono (Ed.), *Language acquisition and the second/foreign language classroom* (pp. 179–211). SEAMEO Regional Language Centre.
<https://files.eric.ed.gov/fulltext/ED367161.pdf#page=191>

- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.
- Ellis, R. (2018). *Reflections on task-based language teaching*. Multilingual Matters.
<https://doi.org/10.21832/ELLIS0131>
- Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, 51(2), 281–318. <https://doi.org/10.1111/1467-9922.00156>
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28(2).
<https://doi.org/10.1017/S0272263106060141>
- Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, 28(4), 575–600.
<https://doi.org/10.1017/S027226310606027X>
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.
- Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning*, 44(3), 449–491.
<https://doi.org/10.1111/j.1467-1770.1994.tb01114.x>
- Felix-Brasdefer, J. C. (2015). *The language of service encounters: A pragmatic-discursive approach*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139565431>
- Felix-Brasdefer, J. C. (2019). Speech acts in interaction. Negotiating joint action in a second language. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics*. Routledge. <https://doi.org/10.4324/9781351164085-2>
- Fernández-García, M., & Arbelaiz, A. M. (2003). Learners' interactions: A comparison of oral and computer-assisted written conversations. *ReCALL*, 15(1), 113–136.
<https://doi.org/10.1017/S0958344003000910>
- Fitze, M. (2006). Discourse and participation in ESL face-to-face and written electronic conferences. *Language Learning & Technology*, 10(1), 67–86.
<https://doi.org/10.125/44047>

- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18(3), 299–323.
<https://doi.org/10.1017/S0272263100015047>
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375.
<https://doi.org/10.1093/applin/21.3.354>
- Freenode. (n.d.). *Freenode IRC webchat* [Computer software].
<https://webchat.freenode.net/>
- Fryer, L., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, 75, 461–468.
<https://doi.org/10.1016/j.chb.2017.05.045>
- Fryer, L., Coniam, D., Carpenter, R., & Lăpuşneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology*, 24(2), 8–22.
<https://doi.org/10125/44719>
- Gass, S. M. (2015). Comprehensible input and output in classroom interaction. In N. Markee (Ed.), *The handbook of classroom discourse and interaction* (pp. 182–197). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118531242.ch11>
- Gass, S. M., & Mackey, A. (2007). Input, interaction and output in SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 175–199). Lawrence Erlbaum Associates.
- Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315813349>
- Gass, S. M., & Varonis, E. M. (1985). Variation in native speaker speech modification to non-native speakers. *Studies in Second Language Acquisition*, 7(1), 37–57.
- Gass, S. M., & Varonis, E. M. (1994). Input, interaction, and second language production. *Studies in Second Language Acquisition*, 16(3), 283–302.
<https://doi.org/10.1017/S0272263100013097>

- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, 34(3), 445-474.
<https://doi.org/10.1017/S0272263112000149>
- Goo, J. (2020). Research on the role of recasts in L2 learning. *Language Teaching*, 53(3), 289-315. <https://doi.org/10.1017/S026144482000004X>
- Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition*, 35(1), 127-165. <https://doi.org/10.1017/S02722631120000708>
- Gurzynski-Weiss, L., & Baralt, M. (2014). Exploring learner perception and use of task-based interactional feedback in FTF and CMC modes. *Studies in Second Language Acquisition*, 36(1), 1-37. <https://doi.org/10.1017/S02722631130000363>
- Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, 29(4), 597-618.
<https://doi.org/10.1093/applin/amno10>
- Holec, H. (1981). *Autonomy and foreign language learning*. Pergamon Press.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
<https://doi.org/10.1093/applin/ampo48>
- Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Language learning & language teaching* (Vol. 32, pp. 1-20). John Benjamins Publishing Company. <https://doi.org/10.1075/lllt.32.01hou>
- Hsu, H.-C. (2017). The effect of task planning on L2 performance and L2 development in text-based synchronous computer-mediated communication. *Applied Linguistics*, 38(3), 359-385. <https://doi.org/10.1093/applin/amvo32>
- Huang, J. (2005). Teacher autonomy in language learning: A review of the research. In *Research studies in education* (Vol. 3, pp. 203-218). Faculty of Education, University of Hong Kong.

- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning – Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237–257. <https://doi.org/10.1111/jcal.12610>
- Hulstijn, J. H., Schoonen, R., de Jong, N. H., Steinel, M. P., & Florijn, A. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR). *Language Testing*, 29(2), 203–221. <https://doi.org/10.1177/0265532211419826>
- Hyland, K. (2018). Genre and discourse analysis in language for specific purposes. In C. A. Chapelle (Ed.), *The encyclopaedia of applied linguistics* (pp. 1–8). <https://doi.org/10.1002/9781405198431.wbealo452.pub2>
- Hymes, D. (1972). Sociolinguistics. In J. Pride & J. Holmes (Eds.), *On communicative competence* (pp. 269–285). Penguin Books.
- Hyon, S. (2018). *Introducing genre and English for specific purposes* (1st ed.). Routledge. <https://doi.org/10.4324/9781315761152>
- IBM SPSS Statistics (Version 27). (2020). [Computer software]. IBM Corp.
- ImmerseMe. (2021). *ImmerseMe* [Computer software]. ImmerseMe Ltd. <https://immerseme.co/>
- Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, 24(2), 168–196. <https://doi.org/10.1093/applin/24.2.168>
- Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition*, 21(3), 421–452. <https://doi.org/10.1017/S0272263199003034>
- Jiang, R. (2022). How does artificial intelligence empower EFL teaching and learning nowadays? A review on artificial intelligence in the EFL context. *Frontiers in Psychology*, 13, 1049401. <https://doi.org/10.3389/fpsyg.2022.1049401>

- Jonsson, E. (2015). *Conversational writing: A multidimensional study of synchronous and supersynchronous computer-mediated communication* (Vol. 16). Peter Lang Publishing Group.
- Junquiera, L., & Kim, Y. (2013). Exploring the relationship between training, beliefs, and teachers' corrective feedback practices: A case study of a novice and an experienced ESL teacher. *The Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, 69(2), 181–206.
- Jurafsky, D., & Martin, J. H. (2023). Chatbots and dialogue Systems. In *Speech and language processing* (3rd ed. Draft, pp. 1–38). Self-published draft.
<https://web.stanford.edu/~jurafsky/slp3/15.pdf>
- Keck, C., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). 3. Investigating the empirical link between task-based interaction and acquisition. A meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 91–131). John Benjamins Publishing Company.
<https://doi.org/10.1075/llt.13.08kec>
- Keller, S., & Reintjes, C. (2016). Aufgaben als Schlüssel zur Kompetenz: Eine Einleitung. In S. Keller & C. Reintjes (Eds.), *Aufgaben als Schlüssel zur Kompetenz* (pp. 15–26). Waxmann. <https://orfee.hepl.ch/handle/20.500.12162/954>
- Kern, R. G. (1995). Restructuring classroom interaction with networked computers: Effects on quantity and characteristics of language production. *The Modern Language Journal*, 79(4), 457–476. <https://doi.org/10.1111/j.1540-4781.1995.tb05445.x>
- Kern, R., & Warschauer, M. (2000). Introduction: Theory and practice of network-based language teaching. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching. Concepts and practice* (pp. 1374–1385). Cambridge University Press.
https://doi.org/10.1007/978-0-387-30424-3_105
- Kim, E. Y. A., & Brown, L. (2014). Negotiating pragmatic competence in computer mediated communication: The case of Korean address terms. *CALICO Journal*, 31(3), 264–284. <https://doi.org/10.1139/cj.31.3.264-284>

- Kim, H.-S., Cha, Y., & Kim, N.-Y. (2020). Impact of mobile interactions with AI on writing performance. *Modern English Education*, 21(2), 1–13.
<https://doi.org/10.18095/meeso.2020.21.2.1>
- Kim, H.-S., Kim, N. Y., & Cha, Y. (2021). Is it beneficial to use AI chatbots to improve learners' speaking performance? *Journal of Asia TEFL*, 18(1), 161–178. Scopus.
<https://doi.org/10.18823/asiatefl.2021.18.1.10.161>
- Kim, N.-Y. (2017a). Effects of different types of chatbots on EFL learners' speaking competence and learner perception. *Cross-Cultural Studies*, 48, 223–252.
<https://doi.org/10.21049/ccs.2017.48..223>
- Kim, N.-Y. (2017b). Effects of types of voice-based chat on EFL students' negotiation of meaning according to proficiency levels. *English Teaching*, 72(1), 159–181.
<https://doi.org/10.15858/engtea.72.1.201703.159>
- Kim, Y. (2015). The role of tasks as vehicles for language learning in classroom interaction. In N. Markee (Ed.), *The handbook of classroom discourse and interaction* (pp. 163–181). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118531242.ch10>
- Kim, Y. (2017). Cognitive-interactionist approaches to L2 instruction. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (1st ed., pp. 126–145). Routledge. <https://doi.org/10.4324/9781315676968>
- Kim, Y., Jung, Y., & Skalicky, S. (2019). Linguistic alignment, learner characteristics, and the production of stranded prepositions in relative clauses: Comparing FTF and SCMC contexts. *Studies in Second Language Acquisition*, 41(5), 937–969.
<https://doi.org/10.1017/S0272263119000093>
- Kim, Y., & McDonough, K. (2008). Learners' production of passives during syntactic priming activities. *Applied Linguistics*, 29(1), 149–154.
<https://doi.org/10.1093/applin/amn004>
- Kim, Y., Skalicky, S., & Jung, Y. (2020). The role of linguistic alignment on question development in face-to-face and synchronous computer-mediated communication contexts: A conceptual replication study. *Language Learning*, 70(3), 643–684.
<https://doi.org/10.1111/lang.12393>

- Koester, A. (2010). *Workplace discourse*. Bloomsbury Publishing.
- Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *International Journal of Mobile Learning and Organisation*, 17(1–2), 214–226. Scopus. <https://doi.org/10.1504/ijmlo.2023.128339>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*. <https://doi.org/10.1177/00336882231162868>
- Koizumi, R. (2012). Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction*, 01(1), 60–69. <https://doi.org/10.7820/vli.vol1.1.koizumi>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition* (1st ed.). Pergamon.
- Kuckartz, U. (2014). *Mixed methods: Methodologie, Forschungsdesigns und Analyseverfahren*. Springer VS. <https://doi.org/10.1007/978-3-531-93267-5>
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (4. Auflage). Beltz Juventa. https://www.beltz.de/produkt_detailansicht/42408-qualitative-inhaltsanalyse-methoden-praxis-computerunterstuetzung.html
- Kuckartz, U., & Rädiker, S. (2019). *Analyzing qualitative data with MAXQDA: Text, audio, and video*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-15671-8>
- Kuiken, F., & Vedder, I. (2018). Chapter 11. Assessing functional adequacy of L2 performance in a task-based approach. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics* (pp. 266–285). John Benjamins Publishing Company. <https://benjamins.com/catalog/tblt.10.11kui>
- Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 81–100). European Second Language Association. http://eurosla.org/monographs/EM01/81-100Kuiken_et_al.pdf

- Lai, C., & Zhao, Y. (2006). Noticing and text-based chat. *Language Learning & Technology*, 10(3), 19. <https://doi.org/10.125/44077>
- Lamb, T. (2008). Introduction to this volume. In T. Lamb & H. Reinders (Eds.), *Learner and teacher autonomy. Concepts, realities, and responses* (Vol. 1, pp. 5–14). John Benjamins Publishing Company.
- Lambert, C., Shintani, N., Skehan, P., Ellis, R., & Li, S. (2019). 2. Cognitive-interactionist perspectives. In C. A. Chapelle & S. Hunston (Eds.), *Task-based language teaching: Theory and practice* (pp. 29–63). Cambridge University Press. <https://doi.org/10.1017/9781108643689.006>
- LanguageTooler GmbH. (2019, January 26). *LanguageTool*. <https://github.com/language-tool-org/language-tool> (Original work published 2013)
- Lantolf, J. P. (2011). 4. Sociocultural theory. A dialectical approach to L2 research. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 57–72). Routledge. <https://doi.org/10.4324/9780203808184.ch4>
- Lee, L. (2001). Online interaction: Negotiation of meaning and strategies used among learners of Spanish. *ReCALL*, 13(2), 232–244. <https://doi.org/10.1017/S0958344001000829a>
- Lee, R. (n.d.). *ESL Robot—English Tutor* (2.0) [Computer software]. rong-chang ESL, Inc. https://www.rong-chang.com/tutor_mike.htm
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309–365. <https://doi.org/10.1111/j.1467-9922.2010.00561.x>
- Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, 97(3), 634–654. <https://doi.org/10.1111/j.1540-4781.2013.12030.x>
- Li, S. (2014). Oral corrective feedback. *ELT Journal*, 68(2), 196–198. <https://doi.org/10.1093/elt/cct076>
- Li, S., & Vuono, A. (2019). Twenty-five years of research on oral and written corrective feedback. *System*, 84, 93–109. <https://doi.org/10.1016/j.system.2019.05.006>

- Lin, W.-C., Huang, H.-T., & Liou, H.-C. (2013). The effects of text-based SCMC on SLA: A meta analysis. *Language Learning & Technology*, 17(2), 123–142.
<https://doi.org/10.125/44327>
- Loewen, S. (2011). The role of feedback. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition*. Routledge.
<https://doi.org/10.4324/9780203808184.ch2>
- Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, 19(1), 1–14.
<https://doi.org/10.1080/09588220600803311>
- Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 361–377). Oxford University Press.
- Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 Classroom: Characteristics, explicitness, and effectiveness. *The Modern Language Journal*, 90(4), 536–556.
<https://doi.org/10.1111/j.1540-4781.2006.00465.x>
- Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, 51(3), 285–329. <https://doi.org/10.1017/S02614444818000125>
- Loewen, S., & Wolff, D. (2016). 6. Peer interaction in F2F and CMC contexts. In M. Sato & S. Ballinger (Eds.), *Peer interaction and second language learning: Pedagogical potential and research agenda* (pp. 163–184). John Benjamins Publishing Company.
<https://benjamins.com/catalog/llt.45.07loe>
- Long, M. H. (1981). Input, interaction, and second-language acquisition. *Annals of the New York Academy of Sciences*, 379(1), 259–278. <https://doi.org/10.1111/j.1749-6632.1981.tb42014.x>
- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4(2), 126–141.
<https://doi.org/10.1093/applin/4.2.126>
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. Ginsberg, & C. Kramersch (Eds.), *Foreign language research in cross-*

- cultural perspective* (pp. 39–52). John Benjamins Publishing Company.
<https://doi.org/10.1075/sibil.2.07lon>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (Vol. 2, pp. 413–468). Academic Press. <https://doi.org/10.1016/b978-012589042-7/50015-3>
- Loschky, L. (1994). Comprehensible input and second language acquisition. *Studies in Second Language Acquisition*, 16(3), 303–322.
<https://doi.org/10.1017/S0272263100013103>
- Lotze, N. (2016). *Chatbots: Eine linguistische Analyse* (1. Auflage). Frankfurt am Main.
- Louis, K., & Cocquio, N. (2019). *Chatbots in English classrooms* [Bachelor's thesis, University of Applied Sciences and Arts Northwestern Switzerland (FHNW)].
<https://webo.fhnw.ch/ht/informatik/i/studierendenprojekte/2018/bachelorarbeiten/cocquio-louis/index.html>
- Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, 59(2), 453–498. <https://doi.org/10.1111/j.1467-9922.2009.00512.x>
- Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, 28(2), 269–300.
<https://doi.org/10.1017/S0272263106060128>
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19(1), 37–66. <https://doi.org/10.1017/S0272263197001034>
- Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32(2), 265–302.
<https://doi.org/10.1017/S0272263109990520>
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46(1), 1–40.
<https://doi.org/10.1017/S0261444812000365>
- Macaro, E. (2008). The shifting dimensions of language learner autonomy. In T. Lamb & H. Reinders (Eds.), *Learner and teacher autonomy. Concepts, realities, and responses* (Vol. 1, pp. 47–62). John Benjamins Publishing Company.

- MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, 44(2), 283–305. <https://doi.org/10.1111/j.1467-1770.1994.tb01103.x>
- Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21(4), 557–587. <https://doi.org/10.1017/S0272263199004027>
- Mackey, A. (2007). The role of conversational interaction in second language acquisition. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 1–26). Oxford University Press.
- Mackey, A., Abbuhl, R., & Gass, S. M. (2011). Interactionist approach. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (1st ed., pp. 7–23). Routledge. <https://doi.org/10.4324/9780203808184.ch1>
- Mackey, A. G., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 407–453). Oxford University Press.
- Mackey, A., & Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, 62(3), 704–740. <https://doi.org/10.1111/j.1467-9922.2011.00649.x>
- Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan UK. <https://doi.org/10.1057/9780230511804>
- Marsden, E., Altmann, G., & Claire, M. S. (2013). Priming of verb inflections in L1 and L2 French: A comparison of ‘redundant’ versus ‘non-redundant’ training conditions. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 51(3), 271–298. <https://doi.org/10.1515/iral-2013-0012>
- Mattermost (5.8). (2019). [Computer software]. Mattermost Inc. <https://mattermost.com/>
- Mayo, M. del P. G., & Soler, E. A. (2013). Negotiated input and output/interaction. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 209–229). <https://doi.org/10.1017/CBO9781139051729.014>

- McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* [Doctoral dissertation, University of Memphis]. The University of Memphis ProQuest Dissertations Publishing.
- McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392. <https://doi.org/10.3758/BRM.42.2.381>
- McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, 27(1), 79–103. <https://doi.org/10.1017/S0272263105050047>
- McDonough, K. (2006). Interaction and syntactic priming: English L2 speakers' production of dative constructions. *Studies in Second Language Acquisition*, 28(2), 179–207. <https://doi.org/10.1017/S0272263106060098>
- McDonough, K. (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 323–338). Oxford University Press.
- McDonough, K., & Chaikitmongkol, W. (2011). Collaborative syntactic priming activities and EFL learners' production of wh-questions. *Canadian Modern Language Review*, 66(6), 817–841. <https://doi.org/10.3138/cmlr.66.6.817>
- McDonough, K., & Mackey, A. (2008). Syntactic priming and ESL question development. *Studies in Second Language Acquisition*, 30(01), 31–47. <https://doi.org/10.1017/S0272263108080029>
- McDonough, K., Neumann, H., & Trofimovich, P. (2015). Eliciting production of L2 target structures through priming activities. *Canadian Modern Language Review*, 71(1), 79–95. <https://doi.org/10.3138/cmlr.2023>
- McHugh, M. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica: Časopis Hrvatskoga Društva Medicinskih Biokemičara/HDMB*, 22, 276–282. <https://doi.org/10.11613/BM.2012.031>

- McNamara, D., Graesser, A. C., & Louwerse, M. M. (2003). *Coh-Metrix* (3.0) [Computer software]. <http://141.225.61.35/cohmetrix2017>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511894664>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 115:1-115:35.
<https://doi.org/10.1145/3457607>
- Melissourgou, M. N., & Frantzi, K. T. (2017). Genre identification based on SFL principles: The representation of text types and genres in English language teaching material. *Corpus Pragmatics*, 1(4), 373-392. <https://doi.org/10.1007/s41701-017-0013-z>
- Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (1st ed., pp. 50-68). Routledge. <https://doi.org/10.4324/9781315676968-4>
- Michel, M., & Cappellini, M. (2019). Alignment during synchronous video versus written chat L2 interactions: A methodological exploration. *Annual Review of Applied Linguistics*, 39, 189-216. <https://doi.org/10.1017/S0267190519000072>
- Michel, M., & O'Rourke, B. (2019). What drives alignment during text chat with a peer vs. a tutor? Insights from cued interviews and eye-tracking. *System*, 83, 50-63.
<https://doi.org/10.1016/j.system.2019.02.009>
- Moussalli, S., & Cardoso, W. (2016). Are commercial 'personal robots' ready for language learning? Focus on second language speech. In S. Papadima-Sophocloeus, L. Bradley, & S. Thouësny (Eds.), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 325-329). Research-publishing.net.
<https://doi.org/10.14705/rpnet.2016.eurocall2016.583>
- Nagata, R., Hashiguchi, T., & Sadoun, D. (2020). Is the simplest chatbot effective in English writing learning assistance? In L. Nguyen, X. Phan, K. Hasida, & S. Tojo (Eds.), *Communications in computer and information science* (Vol. 1215, pp. 245-256). Springer. https://doi.org/10.1007/978-981-15-6168-9_21

- Nakahama, Y., Tyler, A., & van Lier, L. (2001). Negotiation of meaning in conversational and information gap activities: A comparative discourse analysis. *TESOL Quarterly*, 35(3), 377–405.
- Nassaji, H. (2011). Immediate learner repair and its relationship with learning targeted forms in dyadic interaction. *System*, 39(1), 17–29.
<https://doi.org/10.1016/j.system.2011.01.016>
- Nobuyoshi, J., & Ellis, R. (1993). Focused communication tasks and second language acquisition. *English Language Teaching Journal*, 47(3), 203–210.
<https://doi.org/10.1093/elt/47.3.203>
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578.
<https://doi.org/10.1093/applin/ampo44>
- OCam (4.6.5). (n.d.). [Computer software]. OhSoft.
<http://ohsoft.net/eng/ocam/intro.php?cate=1002>
- O'Rourke, B., & Stickler, U. (2017). Synchronous communication technologies for language learning: Promise and challenges in research and pedagogy. *Language Learning in Higher Education*, 7(1), 1–20. <https://doi.org/10.1515/cercles-2017-0009>
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590–601. <https://doi.org/10.1093/applin/ampo45>
- Paltridge, B. (2014). Genre and second language academic writing. *Language Teaching*, 47, 303–318. <https://doi.org/10.1017/S0261444814000068>
- Payne, J. S., & Ross, B. M. (2005). Synchronous CMC, working memory, and L2 oral proficiency development. *Language Learning & Technology*, 9(3), 35–54.
<https://doi.org/10.125/44031>
- Payne, J. S., & Whitney, P. J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal*, 20(1), 7–32. <https://doi.org/10.1558/cj.v20i1.7-32>
- Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In M. Warschauer & R. Kern (Eds.), *Network-based*

- language teaching: Concepts and practice* (pp. 59–86). Cambridge University Press.
<https://doi.org/10.1017/cb09781139524735.006>
- PH Zürich. (n.d.). *Towards a voice-based chatbot for language learners (ChaLL)*. PH Zürich. <https://phzh.ch/de/Forschung/projektdatenbank/projektdetail/?id=379>
- Pica, T. (1996). The essential role of negotiation in the communicative classroom. *JALT Journal*, 18(2), 241–268. <https://jalt-publications.org/jj/articles/2727-essential-role-negotiation-communicative-classroom>
- Pica, T., Kanagy, R., & Falodun, J. (1993). Chapter 9. Choosing and using communication tasks for second language instruction. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (Vol. 93, pp. 9–34). Multilingual Matters. <https://doi.org/10.1075/tblt.1.11cho>
- Pica, T., Lincoln-Porter, F., Paninos, D., & Linnell, J. (1996). Language learners' interaction: How does it address the input, output, and feedback needs of L2 learners? *TESOL Quarterly*, 30(1), 59–84. <https://doi.org/10.2307/3587607>
- Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly*, 21(4), 737. <https://doi.org/10.2307/3586992>
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134(3), 427–459. <https://doi.org/10.1037/0033-2909.134.3.427>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190.
<https://doi.org/10.1017/S0140525X04000056>
- Poarch, G. (2023, June 9). *Multilingualism and cognition—Implications for education*. Evidence-based foreign language learning – Multilingualism in education 2023, Congressi Stefano Franscini Monte Verità, Ascona.
- Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, 26, 10–27. <https://doi.org/10.1016/j.jslw.2014.09.003>
- Quazel. (2023). *Quazel* [Computer software]. Quazel. <https://www.quazel.com/>
- Rasa Technologies GmbH. (n.d.). *Rasa: Open source conversational AI*. <https://rasa.com/>

- Revesz, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37(6), 828–848. <https://doi.org/10.1093/applin/amu069>
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge University Press.
- Rijlaarsdam, G., Couzijn, M., Janssen, T., Braaksma, M., & Kieft, M. (2006). Writing experiment manuals in science education: The impact of writing, genre, and audience. *International Journal of Science Education*, 28(2–3), 203–233. <https://doi.org/10.1080/09500690500336932>
- Rosenthal-von der Pütten, A. M., Straßmann, C., & Krämer, N. C. (2016). Robots or agents – neither helps you more or less during second language acquisition. In D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, & A. Leuski (Eds.), *Intelligent virtual agents* (pp. 256–268). Springer International Publishing. https://doi.org/10.1007/978-3-319-47665-0_23
- Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience*, 11(2), 88–95. <https://doi.org/10.1080/21507740.2020.1740350>
- Sample, E., & Michel, M. (2014). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal*, 31(8), 23–23. <https://doi.org/10.18806/tesl.v31i0.1185>
- Satar, M., & Qinghua, Y. (2020). English as a foreign language learner interactions with chatbots: Negotiations for meaning. *International Online Journal of Education and Teaching (IOJET)*, 7(2), 390–410. <https://iojet.org/index.php/IOJET/article/view/707>
- Sato, M., & Lyster, R. (2012). Peer interaction and corrective feedback for accuracy and fluency development: Monitoring, practice, and proceduralization. *Studies in Second Language Acquisition*, 34(4), 591–626. <https://doi.org/10.1017/S0272263112000356>
- Sato, R. (2012). Self-initiated self-repair attempts by Japanese high school learners while speaking English. *BRAIN. Broad Research in Artificial Intelligence and*

- Neuroscience*, 3(2), 17–28.
<https://lumenpublishing.com/journals/index.php/brain/article/view/1911>
- Sato, R., & Takatsuka, S. (2016). The occurrence and the success rate of self-initiated self-repair. *The Electronic Journal for English as a Second Language*, 20(1), 1–15.
<http://tesl-ej.org/pdf/ej77/a4.pdf>
- Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, 13(1). <https://doi.org/10.10125/44170>
- Sauro, S. (2011). SCMC for SLA: A research synthesis. *CALICO Journal*, 28(2), 369–391.
<https://doi.org/10.11139/cj.28.2.369-391>
- Sauro, S. (2012). L2 performance in text-chat and spoken discourse. *System*, 40(3), 335–348. <https://doi.org/10.1016/j.system.2012.08.001>
- Sauro, S. J. (2013). 7. The cyber language exchange: Cross-national computer-mediated interaction. In K. McDonough & A. Mackey (Eds.), *Language learning & language teaching* (Vol. 34, pp. 129–146). John Benjamins Publishing Company.
<https://doi.org/10.1075/llt.34.1ch7>
- Sauro, S., & Smith, B. (2010). Investigating L2 performance in text chat. *Applied Linguistics*, 31(4), 554–577. <https://doi.org/10.1093/applin/amq007>
- SBFI. (2012). *Verordnung des SBFI über die berufliche Grundbildung Kauffrau/Kaufmann mit eidgenössischem Fähigkeitszeugnis (EFZ)*.
<https://www.skkab.ch/download/bildungsverordnung-kauffrau-kaufmann-efz-2/?wpdmdl=448&refresh=64b683f017dd81689682928>
- SBFI. (2021). *Verordnung des SBFI über die berufliche Grundbildung Kauffrau/Kaufmann mit eidgenössischem Fähigkeitszeugnis (EFZ)*.
<https://www.becc.admin.ch/becc/public/bvz/beruf/download/11540>
- SBFI. (2022). *Berufsbildung in der Schweiz – Fakten und Zahlen 2022*.
<https://www.sbf.admin.ch/sbf/de/home/dienstleistungen/publikationen/publikationsdatenbank/berufsbildung-in-der-schweiz.html>
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–369). Newbury House Publishers.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schocker, M., & Müller-Hartmann, A. (2016). Professionalisierung durch forschendes Erfahrungslernen. Lehrkompetenzen im aufgabenorientierten Englischunterricht mit Hilfe von Unterrichtsvideos entwickeln. In S. Keller & C. Reintjes (Eds.), *Aufgaben als Schlüssel zur Kompetenz. Didaktische Herausforderungen, wissenschaftliche Zugänge und empirische Befunde* (pp. 53–72). Waxmann.
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8(3), 263–300. <https://doi.org/10.1191/1362168804lr1460a>
- Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research*, 10(4), 361–392. <https://doi.org/10.1191/1362168806lr2030a>
- Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language Learning*, 58(4), 835–874. <https://doi.org/10.1111/j.1467-9922.2008.00480.x>
- Sheen, Y., & Ellis, R. (2013). Corrective feedback in language teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 593–610). Routledge. <https://doi.org/10.4324/9780203836507.ch36>
- Shehadeh, A. (2001). Self- and other-initiated modified output during task-based interaction. *TESOL Quarterly*, 35(3), 433–457. <https://doi.org/10.2307/3588030>
- Sinclair, A., McCurdy, K., Lucas, C. G., Lopez, A., & Gašević, D. (2019). Tutorbot corpus: Evidence of human-agent verbal alignment in second language learner dialogues. In C. F. Lynch, A. Merceron, M. Desmarais, & R. Nkambou (Eds.), *Proceedings of the 12th International Conference on Educational Data Mining* (pp. 414–419). <https://educationaldatamining.org/edm2019/proceedings/>

- Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, 18, 268–286. <https://doi.org/10.1017/S0267190500003585>
- Skehan, P. (2003). Focus on form, tasks, and technology. *Computer Assisted Language Learning*, 16(5), 391–411. <https://doi.org/10.1076/call.16.5.391.29489>
- SKKAB. (2020). *Bildungsplan Kauffrau/Kaufmann EFZ vom 26. September 2011 für die betrieblich organisierte Grundbildung (Stand am 1. Januar 2020)*. Schweizerische Konferenz der kaufmännischen Ausbildungs- und Prüfungsbranchen. <https://www.skkab.ch/download/bildungsplan-fuer-die-betrieblich-organisierte-grundbildung/?wpdmdl=447&refresh=64b6ba2998bbe1689696809>
- SKKAB. (2021a). *Bildungsplan zur Verordnung des SBFJ vom 16. August 2021 über die berufliche Grundbildung für Kauffrau/Kaufmann mit eidgenössischem Fähigkeitszeugnis (EFZ) vom 24. Juni 2021 (Stand am 1. Juni 2023)*. Schweizerische Konferenz der kaufmännischen Ausbildungs- und Prüfungsbranchen. https://www.skkab.ch/download/bildungsplan/?wpdmdl=6259&refresh=64b6beff47bda1689698047&ind=1685604344061&filename=DE_BiPla_SKKAB_inkl_Branchenspezifika_Version-01.06.2023.pdf
- SKKAB. (2021b). *Fremdsprachenkonzept Kauffrau/Kaufmann EFZ*. Schweizerische Konferenz der kaufmännischen Ausbildungs- und Prüfungsbranchen. <https://www.skkab.ch/aktuell/fit-fuer-die-schweiz-bereit-fuer-die-welt/>
- Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal*, 87(1), 38–57. <https://doi.org/10.1111/1540-4781.00177>
- Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition*, 26(3), 365–398. <https://doi.org/10.1017/S027226310426301X>
- Smith, B. (2005). The relationship between negotiated interaction, learner uptake, and lexical acquisition in task-based computer-mediated communication. *TESOL Quarterly*, 39(1), 33–58. <https://doi.org/10.2307/3588451>

- Smith, B., & Michel, M. (2017). Measuring lexical alignment during L2 chat interaction: An eye tracking study. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Saliency in second language acquisition* (1st ed., pp. 244–268). Routledge.
- Smith, B., & Renaud, C. (2013). 8. Using eye tracking as a measure of foreign language learners' noticing of recasts during computer-mediated writing conferences. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational contexts* (pp. 147–166). John Benjamins Publishing Company.
<https://doi.org/10.1075/lllt.34.12ch8>
- Sotillo, S. (2005). Corrective feedback via instant messenger learning activities in NS-NNS and NNS-NNS dyads. *CALICO Journal*, 22(3), 467–496.
<https://doi.org/10.1558/cj.v22i3.467-496>
- Sotillo, S. M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*, 4(1), 82–119.
<https://doi.org/10125/25088>
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation*, 9(4). <https://doi.org/10.7275/96JP-XZ07>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125–144). Oxford University Press.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 495–508). Routledge. <https://doi.org/10.4324/9781410612700-34>
- Swales, J. (2009). Chapter 3: The concept of task. In K. Van den Branden, M. Bygate, & J. M. Norris (Eds.), *Task-based language teaching: A reader* (pp. 41–56). John Benjamins Publishing.

- Sydorenko, T. (2015). The use of computer-delivered structured tasks in pragmatic instruction: An exploratory study. *Intercultural Pragmatics*, 12(3), 333–362. <https://doi.org/10.1515/ip-2015-0017>
- Sydorenko, T., Daurio, P., & Thorne, S. L. (2018). Refining pragmatically-appropriate oral communication via computer-simulated conversations. *Computer Assisted Language Learning*, 31(1–2), 157–180. <https://doi.org/10.1080/09588221.2017.1394326>
- Sydorenko, T., Smits, T. F. H., Evanini, K., & Ramanarayanan, V. (2019). Simulated speaking environments for language learning: Insights from three cases. *Computer Assisted Language Learning*, 32(1–2), 17–48. <https://doi.org/10.1080/09588221.2018.1466811>
- Sykes, J. M. (2005). Synchronous CMC and pragmatic development: Effects of oral and written chat. *CALICO Journal*, 22(3), 399–431. <https://doi.org/10.1558/cj.v22i3.399-431>
- Tai, T.-Y., & Chen, H. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485–1502. <https://doi.org/10.1080/10494820.2020.1841801>
- Tavakoli, P., & Skehan, P. (2005). 9. Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). John Benjamins Publishing Company.
- Templin, M. C. (1957). *Certain language skills in children: Their development and interrelationships*. University of Minnesota Press. <https://doi.org/10.5749/j.ctttv2st>
- Thonhauser, I. (2016). Was macht gute Aufgaben für den Fremdsprachenunterricht aus? Charakteristik guter Aufgaben und Einsichten aus der Unterrichtsbeobachtung. In S. Keller & C. Reintjes (Eds.), *Aufgaben als Schlüssel zur Kompetenz. Didaktische Herausforderungen, wissenschaftliche Zugänge und empirische Befunde* (pp. 179–196). Waxmann. <https://orfee.hepl.ch/handle/20.500.12162/954>
- Timpe-Laughlin, V., & Dombi, J. (2020). Exploring L2 learners' request behavior in a multi-turn conversation with a fully automated agent. *Intercultural Pragmatics*, 17(2), 221–257. <https://doi.org/10.1515/ip-2020-0010>

- Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, 4(4), 156–163. [https://doi.org/10.1016/S1364-6613\(00\)01462-5](https://doi.org/10.1016/S1364-6613(00)01462-5)
- Trofimovich, P., McDonough, K., & Foote, J. A. (2014). Interactive alignment of multisyllabic stress patterns in a second language classroom. *TESOL Quarterly*, 48(4), 815–832. <https://doi.org/10.1002/tesq.156>
- Tudini, V. (2007). Negotiation and intercultural learning in Italian native speaker chat rooms. *The Modern Language Journal*, 91(4), 577–601. <https://doi.org/10.1111/j.1540-4781.2007.00624.x>
- Turner, C. E., & Upshur, J. A. (1995). Some effects of task type on the relation between communicative effectiveness and grammatical accuracy in intensive ESL classes. *TESL Canada Journal*, 12(2), 18–31. <https://doi.org/10.18806/tesl.v12i2.650>
- Unipark (10.9). (2019). [Computer software]. Questback Deutschland GmbH. <https://www.unipark.com/>
- University of Cambridge. Local Examinations Syndicate. (2001). *Quick placement test: Paper and pen version pack*. Oxford University Press.
- Uzum, B. (2010). An investigation of alignment in CMC from a sociocognitive perspective. *CALICO Journal*, 28(1), 135–155. <https://doi.org/10.1139/cj.28.1.135-155>
- Varonis, E. M., & Gass, S. (1985). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, 6(1), 71–90. <https://doi.org/10.1093/applin/6.1.71>
- VERBI Software. (2020). *MAXQDA (Version 2020)* [MAXQDA 2020]. VERBI Software. <https://www.maxqda.com>
- Vindayer, C., & Hodel, H.-P. (2007). *Der Fremdsprachenunterricht auf Sekundarstufe II: Bestandesaufnahme*. Schweizerische Konferenz der kantonalen Erziehungsdirektoren. https://edudoc.ch/record/88590/files/beili_bestandesaufnahme_d.pdf?ln=de

- Ware, P. (2005). "Missed" communication in online communication: Tensions in a German-American telecollaboration. *Language Learning & Technology*, 9(2), 64–89. <https://doi.org/10.125/44020>
- Warschauer, M. (1996a). Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal*, 13(2–3), 7–26. <https://doi.org/10.1558/cj.v13i2-3.7-26>
- Warschauer, M. (1996b). Motivational aspects of using computers for writing and communication. *Telecollaboration in Foreign Language Learning*, 29–46. <http://hdl.handle.net/10125/8946>
- Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *The Modern Language Journal*, 81(4), 470–481. <https://doi.org/10.1111/j.1540-4781.1997.tb05514.x>
- Whitby, N. (2013). *Business Benchmark. Pre-intermediate to intermediate. Student's Book*. (2nd ed.). Cambridge University Press.
- White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Interlanguage Studies Bulletin (Utrecht)*, 7(2), 133–161. <https://doi.org/10.1177/026765839100700205>
- Wilske, S. (2015). *Form and meaning in dialog-based computer-assisted language learning* [Doctoral dissertation, Universität des Saarlandes]. SciDok – Der Wissenschaftsserver der Universität des Saarlandes. 10.22028/D291-23654
- Wilske, S., & Wolska, M. (2011). Meaning versus form in computer-assisted task-based language learning: A case study on the German dative. *Journal for Language Technology and Computational Linguistics*, 26, 23–37. <https://doi.org/10.21248/jlcl.26.2011.134>
- WKS. (n.d.). *Kauffrau/Kaufmann EFZ BGSOL 2.o BGSOL – Begleitetes selbstorganisiertes Lernen*. <https://www.wksbern.ch/de/lernende-lehrbetriebe/kauffrau-kaufmann-efz-bgsol-e-profil>
- Wolska, M., & Wilske, S. (2010a). German subordinate clause word order in dialogue-based CALL. *Proceedings of the International Multiconference on Computer Science*

- and Information Technology (pp. 553–559). IEEE.
<https://doi.org/10.1109/IMCSIT.2010.5679620>
- Wolska, M., & Wilske, S. (2010b, September 22). Form-focused task-oriented dialogues for computer assisted language learning: A pilot study on German dative. *Proceedings of the INTERSPEECH 2010 Workshop on “Second Language Studies: Acquisition, Learning, Education and Technology.”* ISCA. https://www.isca-speech.org/archive/pdfs/l2ws_2010/wolska10_l2ws.pdf
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*.
<https://doi.org/10.1007/s10639-023-11742-4>
- Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners’ acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, 32(2), 235–263.
<https://doi.org/10.1017/S0272263109990519>
- Yilmaz, Y. (2016). The role of exposure condition in the effectiveness of explicit correction. *Studies in Second Language Acquisition*, 38(1), 65–96.
<https://doi.org/10.1017/S0272263115000212>
- Yilmaz, Y., & Yuksel, D. (2011). Effects of communication mode and salience on recasts: A first exposure study. *Language Teaching Research*, 15(4), 457–477.
<https://doi.org/10.1177/1362168811412873>
- Yoshida, R. (2010). How do teachers and learners perceive corrective feedback in the Japanese language classroom? *The Modern Language Journal*, 94(2), 293–314.
<https://doi.org/10.1111/j.1540-4781.2010.01022.x>
- Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, 32(2), 199–225.
<https://doi.org/10.1177/0265532214557113>
- Yuksel, D., & Inan, B. (2014). The effects of communication mode on negotiation of meaning and its noticing. *ReCALL*, 26(3), 333–354.
<https://doi.org/10.1017/S0958344014000147>

- Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *The Modern Language Journal*, 103(4), 763–781.
<https://doi.org/10.1111/modl.12590>
- Ziegler, N. (2016). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*, 38(3), 553–586.
<https://doi.org/10.1017/S027226311500025X>
- Ziegler, N. (2018). Pre-task planning in L2 text-chat: Examining learners' process and performance. *Language Learning & Technology*, 22(3), 193–213.
<https://doi.org/10125/44664>
- Ziegler, N., & Bryfonski, L. (2018). Interaction-driven L2 learning. In *The handbook of advanced proficiency in second language acquisition* (pp. 94–113). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119261650.ch6>
- Ziegler, N., & Mackey, A. (2017). Interactional feedback in synchronous computer-mediated communication. A review of the state of the art. In N. Hossein & K. Eva (Eds.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (1st ed., pp. 80–94). Routledge.
<https://doi.org/10.4324/9781315621432>
- Ziegler, N., & Phung, H. (2019). Technology-mediated task-based interaction: The role of modality. *ITL – International Journal of Applied Linguistics*, 170(2), 251–276.
<https://doi.org/10.1075/itl.19014.zie>
- ZTO Labs. (2021). *Andy* [Computer software]. ZTO Labs.
<https://play.google.com/store/apps/details?id=com.pyankoff.andy>

Appendices

Appendix A Topics and objectives of participating schools' English curriculum

Interner Lehrplan Englisch Kauffrau / Kaufmann E-Profil

10. Inhalte des Englischunterrichts: detaillierter Lehrplan E-Profil

Lehrjahr 1: Semester 1

ZEITRAUM	WOCHEN / LEKTIONEN	LEHRMITTEL / UNITS	THEMEN
August bis Januar	18 / 3	Headway Pre-Intermediate Units 1 – 6 + Present Perfect Tense	<ul style="list-style-type: none"> • Informationen über sich und andere bzw. über den Alltag geben und erfragen • Über vergangene Ereignisse berichten • Gelesenem und Gehörtem die wesentlichen Informationen entnehmen • Einkaufen, Preise und Mengenangaben beherrschen • Gefühle und Pläne mitteilen • Vergleiche anstellen • Personen und Orte beschreiben • Meinungen begründen und austauschen • Kurze Texte, z.B. Briefe oder E-Mails verfassen
Woche 42/43	1 / 3	Lerninsel 1: Telefonieren und Telefonnotizen ausfüllen / Rollenspiele	<ul style="list-style-type: none"> • Wortschatz (Einzelwörter) • spezielle Ausdrücke und Wendungen

Lehrjahr 1: Semester 2

ZEITRAUM	WOCHEN / LEKTIONEN	LEHRMITTEL / UNITS	THEMEN
Februar bis Juli	17 / 3	Business Benchmark Units 1 - 8	<ul style="list-style-type: none"> • Informelle Texte verfassen • Wegbeschreibungen verstehen und geben • Sich über zukünftige Ereignisse und Möglichkeiten austauschen • Entscheidungen diskutieren und begründen • Ratschläge geben • Höflichkeitsfloskeln kennen und verwenden • Texte und Dialoge über Zeitgeschehen verstehen und die wesentlichen Informationen entnehmen und weitergeben
ca. Woche 48: Placement Test for FCE Courses (online auf Campus)		Inhalt Vergleichstest: Headway Units 1-6 Present Perfect tense (Zusatzmaterial) Telephoning Business Vocabulary Units 1-2 from Business Benchmark	
Kalenderwoche 12: Vergleichstest			
Woche 25	1 / 3	Lerninsel 2: Sich vorstellen, Kundinnen / Kunden und Gäste empfangen	<ul style="list-style-type: none"> • Wortschatz (Einzelwörter) • spezielle Ausdrücke und Wendungen • situationsgerechtes Reagieren
Woche 26	1 / 3	Lerninsel 3: Unternehmen, Produkt oder Dienstleistungen präsentieren	<ul style="list-style-type: none"> • Wortschatz (Einzelwörter) • spezielle Ausdrücke und Wendungen • situationsgerechtes Reagieren

Appendix B Communicative functions in coursebook sample texts matching dialog system's actions

Action	Communicative Function
Confirm	I can confirm that the office space includes private bathroom facilities. We would be happy to arrange a viewing of the premises.
Offer additional service	I will be pleased to provide any further information you may require.
Request	I would be grateful if you could fill in the form on the website as soon as possible in order to confirm your booking. If you would like to arrange a consultation, please call us on the above number to fix a date and time. I look forward to hearing from you. Please call our office on 01343 758495 to fix a suitable time. Please could you let us know if you would be willing to contribute? So please could you let me know if your company would like to enter. We would be grateful if you could complete this and return it to us.
Thank	Many thanks for your enquiry about our training courses. Thank you for your enquiry.

Appendix C Task handouts

System group

Versetzen Sie sich in die folgende Situation. Führen Sie den Auftrag mit dem Chatbot durch. Schreiben Sie so, wie Sie es für die Arbeit angemessen finden.

--

Sie sind KV-Lernende/r in einem Betrieb. Ihre Kolleginnen und Kollegen aus der Marketing-Abteilung organisieren am 12. Oktober 2019 eine Firmenveranstaltung in Toronto (Kanada). Sie haben nun den Auftrag, im «One King» Hotel in Toronto einen passenden Raum zu finden.

Der Raum soll:

- Platz haben für mind. 150 Personen
- eine Präsentation (Video, Audio) ermöglichen
- Gespräche in kleineren Gruppen ermöglichen
- eine stimmungsvolle Atmosphäre haben
- den Gästen Eindruck machen, aber nicht zu protzig/übertrieben wirken
- max. 1'500 Franken kosten

Das Hotel «One King» hat drei Räume: Alpha, Beta und Gamma. Kontaktieren Sie das Hotel über ihren Online-Chat und stellen Sie dem Hotel-Chatbot Fragen zu den drei Räumen.

Wenn Sie alle nötigen Informationen haben, buchen Sie den Raum, der Ihrer Meinung nach am besten geeignet ist für die Veranstaltung.

SCMC group (learner/client role)

Versetzen Sie sich in die folgende Situation. Führen Sie den Auftrag über Chat mit einer Kollegin / einem Kollegen aus der Klasse durch. Schreiben Sie so, wie Sie es für die Arbeit angemessen finden.

--

Sie sind KV-Lernende/r in einem Betrieb. Ihre Kolleginnen und Kollegen aus der Marketing-Abteilung organisieren am 12. Oktober 2019 eine Firmenveranstaltung in Toronto (Kanada). Sie haben nun den Auftrag, im «One King» Hotel in Toronto einen passenden Raum zu finden.

Der Raum soll:

- Platz haben für mind. 150 Personen
- eine Präsentation (Video, Audio) ermöglichen
- Gespräche in kleineren Gruppen ermöglichen
- eine stimmungsvolle Atmosphäre haben
- den Gästen Eindruck machen, aber nicht zu protzig/übertrieben wirken
- max. 1'500 Franken kosten

Das Hotel «One King» hat drei Räume: Alpha, Beta und Gamma. Kontaktieren Sie das Hotel über ihren Online-Chat und stellen Sie dem / der Hotel-Mitarbeitenden Fragen zu den drei Räumen.

Wenn Sie alle nötigen Informationen haben, buchen Sie den Raum, der Ihrer Meinung nach am besten geeignet ist für die Veranstaltung.





SCMC group (interlocutor/hotel employee role)

Versetzen Sie sich in die folgende Situation. Führen Sie den Auftrag über Chat mit einer Kollegin / einem Kollegen aus der Klasse durch. Schreiben Sie so, wie Sie es für die Arbeit angemessen finden.

–

Sie arbeiten im «One King» Hotel in Toronto (Kanada). Heute werden Sie über den Online-Chat von einem potentiellen Kunden / einer potentiellen Kundin kontaktiert, der / die einen Ihrer drei Räume mieten möchte.

Beantworten Sie seine / ihre Fragen zu den drei Räumen mithilfe der nachfolgenden Informationen.

Raum:	Alpha	Beta	Gamma
Max. Gäste:	270	130 (Innenbereich) + 60 (Terrasse)	200
Mobiliar:	<ul style="list-style-type: none"> - Bühne - fixe Bestuhlung 	<p>Innenbereich:</p> <ul style="list-style-type: none"> - Bar mit Barhockern - Tische für 4 Personen - Oberlicht <p>Terrasse:</p> <ul style="list-style-type: none"> - Lounge-Ecken mit niedrigen Lounge-Tischen und Sesseln - runde Stehtische mit Barhockern 	<ul style="list-style-type: none"> - runde Tische für 6-8 Personen - Stühle - Mobiliar kann im Raum verschoben werden - original Dekoration aus den 1920er Jahren
Technische Ausstattung:	modernste Multimedialechnik (Grossleinwand, Projektor, Tonanlage) und Lichtanlage	mobile Leinwand, Projektor, Lautsprecher	2 grosse mobile Bildschirme, Lautsprecher
Lage:	im Untergeschoss	oberster Stock mit Aussicht über die Stadt	im 1. Stock
Preis:	1'400 Franken	900 Franken	1'100 Franken
Bilder:		 	

Appendix D Learner questionnaires

Pre-questionnaire

Anzeigeeoptionen

Info: Hier können Sie optional die Anzeigeeoptionen verändern. Wenn Sie eine Sprache auswählen, die keine eigenen Textelemente hat, werden die Textelemente der Standardsprache angezeigt.

Anzeigeeoptionen einstellen:

- Anzeigeeoptionen mit ausdrucken
- Interne Verlinkungen einblenden
- Filter anzeigen
- Trigger anzeigen
- Plauschecks anzeigen
- Randomisierung von Antwortoptionen anschalten
- Pretest-Kommentare anzeigen
- Todos anzeigen

Sprache German ▾

Informationen zur Umfrage Hauptstudie – Pre-Fragebogen für Lernende

Umfrage-Nr.	566454
Autor	Johanna Oeschger
Co-Autoren	
Start	29.04.2019 00:00:00
Ende	10.05.2019 18:00:00

Fragebogen

1 Code

Bitte geben Sie Ihren persönlichen Code ein.

Format: XXX-XXX

2 Willkommen

Liebe Lernende

Bei der heutigen Studie geht es darum, neue Aufgaben für den Englischunterricht in der Praxis zu testen. Mit diesem Fragebogen möchten wir Ihre bisherigen Erfahrungen mit dieser Art von Aufgaben erfassen.

Bitte lesen Sie die Fragen aufmerksam durch und wählen Sie die Antwort, die am besten auf Sie zutrifft. Es gibt keine richtigen oder falschen Antworten. Ihre Angaben sind sehr wichtig für uns, bitte beantworten Sie deshalb wenn möglich alle Fragen.

Das Ausfüllen des Fragebogens dauert etwa 5 Minuten.

Herzlichen Dank für Ihre Teilnahme!

3 Alter

Wie alt sind Sie?

4 Geschlecht

Welches Geschlecht haben Sie?

- Weiblich
- Männlich
- Sonstiges
-

5 Chatten Englisch

Wie oft chatten Sie auf Englisch?

Gemeint ist Chatten über soziale Medien wie z.B. Whatsapp, Snapchat, Instagram Direktnachrichten, Facebook Messenger.

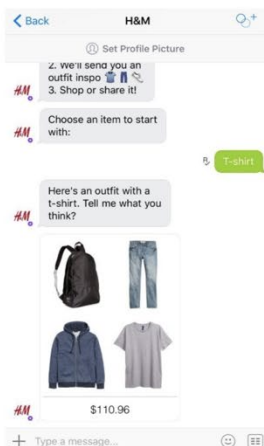
- Nie oder fast nie
- 1-3 Mal im Jahr
- 1-3 Mal pro Monat
- 1-3 Mal in der Woche
- Täglich oder fast täglich
-

6 Dialogsystem

Wie oft chatten oder sprechen Sie mit Robotern oder anderen Geräten?

Gemeint sind:

- *Digitale Assistenten auf dem Handy oder Computer (z.B. Siri, Google Assistant)*
- *Chatbots auf Websites oder in Apps, die z.B. Produkte vorschlagen, Bestellungen ausführen oder Kundenanfragen beantworten*
- *Chatbots zur Unterhaltung (z.B. Cleverbot)*
- *Sprachgesteuerte Geräte (z.B. Amazon Alexa, Google Home)*
- *Roboter*



- Nie oder fast nie
- 1-3 Mal pro Jahr
- 1-3 Mal pro Monat
- 1-3 Mal pro Woche
- Täglich oder fast täglich

Wie oft chatten oder sprechen Sie mit Robotern oder anderen Geräten auf Englisch?

- Nie oder fast nie
- 1-3 Mal pro Jahr
- 1-3 Mal pro Monat
- 1-3 Mal pro Woche
- Täglich oder fast täglich

7 Schreiben EN

Wie oft kommunizieren Sie beruflich mit anderen Leuten über E-Mail oder Briefe auf Englisch?

- Nie oder fast nie
- 1-3 Mal pro Jahr
- 1-3 Mal pro Monat
- 1-3 Mal pro Woche

- Täglich oder fast täglich
-

8 Schreiben EN privat

Wie oft kommunizieren Sie privat mit anderen Leuten über E-Mail auf Englisch?

- Nie oder fast nie
- 1-3 Mal pro Jahr
- 1-3 Mal pro Monat
- 1-3 Mal pro Woche
- Täglich oder fast täglich
-

9 Soziale Medien EN

Wie oft posten oder kommentieren Sie in sozialen Medien (z.B. Instagram, Snapchat Stories, Facebook, Youtube), Blogs oder Foren (z.B. gutefrage, reddit) auf Englisch?

- Nie oder fast nie
- 1-3 Mal pro Jahr
- 1-3 Mal pro Monat
- 1-3 Mal pro Woche
- Täglich oder fast täglich
-

10 Abschlussfrage

Haben Sie Fragen oder Bemerkungen?

11 Final page

Herzlichen Dank für Ihre Teilnahme!

Post-questionnaire

Anzeigeoptionen

Info: Hier können Sie optional die Anzeigeoptionen verändern. Wenn Sie eine Sprache auswählen, die keine eigenen Textelemente hat, werden die Textelemente der Standardsprache angezeigt.

Anzeigeoptionen einstellen:

- Anzeigeoptionen mit ausdrucken
- Interne Verlinkungen einblenden
- Filter anzeigen
- Trigger anzeigen
- Plauschecks anzeigen
- Randomisierung von Antwortoptionen anschalten
- Pretest-Kommentare anzeigen
- Todos anzeigen

Sprache

Informationen zur Umfrage Hauptstudie – Post-Fragebogen für Lernende

Umfrage-Nr.	566473
Autor	Johanna Oeschger
Co-Autoren	
Start	29.04.2019 00:00:00
Ende	10.05.2019 18:00:00

Fragebogen

1 Code

Bitte geben Sie Ihren persönlichen Code ein.

Format: XXX-XXX

2 Willkommen

Liebe Lernende

Bei der heutigen Studie geht es darum, neue Aufgaben für den Englischunterricht in der Praxis zu testen. Mit diesem Fragebogen möchten wir nun Ihr Feedback zu der Aufgabe einholen.

Bitte lesen Sie die Fragen aufmerksam durch und wählen Sie die Antwort, die am besten auf Sie zutrifft. Es gibt keine falschen oder richtigen Antworten. Ihre Angaben sind sehr wichtig für uns, bitte beantworten Sie daher wenn möglich alle Fragen.

Das Ausfüllen des Fragebogens dauert etwa 10 Minuten.

Herzlichen Dank für Ihre Teilnahme!

3 Gern gemacht

Ich habe die Aufgabe gern gemacht.

Trifft nicht zu

-

- Trifft eher nicht zu
 - Teils-teils
 - Trifft eher zu
 - Trifft zu
-

4 Spannend

Ich fand die Aufgabe spannend.

- Trifft nicht zu
 - Trifft eher nicht zu
 - Teils-teils
 - Trifft eher zu
 - Trifft zu
-

5 Nützlich

Die Aufgabe ist nützlich für das Englischlernen.

- Trifft nicht zu
 - Trifft eher nicht zu
 - Teils-teils
 - Trifft eher zu
 - Trifft zu
-

6 Mühe gegeben

Ich habe mir Mühe gegeben.

- Trifft nicht zu
- Trifft eher nicht zu
-

Teils-teils

- Trifft eher zu
 - Trifft zu
-

7 Gut gelungen

Die Aufgabe ist mir gut gelungen.

- Trifft nicht zu
 - Trifft eher nicht zu
 - Teils-teils
 - Trifft eher zu
 - Trifft zu
-

8 Englischkenntnisse

Meine Englischkenntnisse waren ausreichend, um die Aufgabe zu bearbeiten.

- Trifft nicht zu
 - Trifft eher nicht zu
 - Teils-teils
 - Trifft eher zu
 - Trifft zu
-

9 Inhaltliche Bearbeitung

Inhaltlich fiel mir die Aufgabe leicht.

- Trifft nicht zu
- Trifft eher nicht zu
- Teils-teils
-

Trifft eher zu

Trifft zu

10 Technische Schwierigkeiten

Gab es technische Schwierigkeiten (z.B. Internetverbindung, Verzögerung beim Übermitteln der Nachrichten, Tippen usw.), die sich auf das Ausführen der Aufgabe ausgewirkt haben?

Ja

Nein

11.1 Technische Schwierigkeiten

Welche technischen Schwierigkeiten gab es?

12 Gesprächspartner verstanden

Haben Sie die Aussagen Ihres Gesprächspartners verstanden?

Selten verstanden

Manchmal verstanden

Oft verstanden

Immer verstanden

13.1 Gesprächspartner nicht verstanden

Wie gingen Sie vor, wenn Sie Ihren Gesprächspartner nicht verstanden?

Bitte kreuzen Sie alle zutreffenden Antworten an.

Ich sagte dem Gesprächspartner, dass ich ihn nicht verstehe.

Ich fragte beim Gesprächspartner nach, ob ich ihn richtig verstehe.

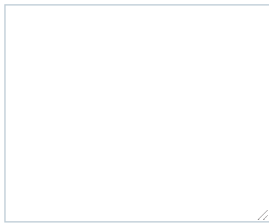
Ich bat den Gesprächspartner, seine Aussage zu erklären.

-

- Ich schaute im Internet nach.
- Ich versuchte, die Bedeutung aus dem Kontext herzuleiten.
- Ich schaute weiter oben im Chat nach.
- Ich wechselte das Thema.
- Ich tat nichts.
- Etwas anderes.

14.1 Anderes Vorgehen wenn nicht verstanden

Sie haben bei der letzten Frage «etwas anderes» ausgewählt. Wie gingen Sie sonst noch vor, wenn Sie Ihren Gesprächspartner nicht verstanden?



15 Vom Gesprächspartner verstanden

Wurden Sie von Ihrem Gesprächspartner verstanden?

- Selten verstanden
- Manchmal verstanden
- Oft verstanden
- Immer verstanden

16.1 Vom Gesprächspartner nicht verstanden

Wie gingen Sie vor, wenn Sie von Ihrem Gesprächspartner nicht verstanden wurden?

Bitte kreuzen Sie alle zutreffenden Antworten an.

- Ich schrieb die Aussage noch einmal, ohne sie zu verändern.
- Ich veränderte die Rechtschreibung.
-

Ich verwendete ein anderes Wort.

- Ich veränderte die Satzstellung.
- Ich schrieb die Aussage ausführlicher und/oder detaillierter.
- Ich schrieb die Aussage weniger ausführlich und/oder weniger detailliert.
- Ich teilte die Aussage auf mehrere Nachrichten auf.
- Ich fasste mehrere Nachrichten in einer zusammen.
- Ich fragte den Gesprächspartner, ob er mich verstanden hatte.
- Ich schrieb auf Deutsch.
- Ich wechselte das Thema.
- Etwas anderes.

17.1 Anderes Vorgehen wenn vom Gesprächspartner nicht verstanden

Sie haben bei der letzten Frage «etwas anderes» ausgewählt. Wie gingen Sie sonst noch vor, wenn Sie von Ihrem Gesprächspartner nicht verstanden wurden?

18 Angemessen behandelt

Fühlten Sie sich von Ihrem Gesprächspartner angemessen behandelt?

- Selten angemessen behandelt
- Manchmal angemessen behandelt
- Oft angemessen behandelt
- Immer angemessen behandelt

19 Eigene Textart

Wie haben Sie geschrieben?

Bitte wählen Sie in jeder Zeile die Beschreibung aus, die besser auf Ihre Schreibart zutrifft.

-

Gespräch mit Arbeitskollegen Kundengespräch

Auf Höflichkeitsformen verzichten Höflichkeitsformen anwenden

Geschäftlich Privat

Auf Rechtschreibung achten Nicht auf Rechtschreibung achten

Business English Alltagsenglisch

Komplexe Sätze Einfache Sätze

Abwechslungsreiche Wörter und Satzstellungen Möglichst einfach verständliche Wörter und Satzstellungen

Formell Informell

Mündlicher Stil Schriftlicher Stil

20 Textart ideal

Wie sollte man Ihrer Meinung nach schreiben bei dieser Aufgabe?

Bitte wählen Sie in jeder Zeile die Beschreibung aus, die besser auf die Schreibart zutrifft.

Mündlicher Stil Schriftlicher Stil

Business English Alltagsenglisch

Formell Informell

Auf Höflichkeitsformen verzichten Höflichkeitsformen anwenden

Auf Rechtschreibung achten Nicht auf Rechtschreibung achten

Abwechslungsreiche Wörter und Satzstellungen Möglichst einfach verständliche Wörter und Satzstellungen

Komplexe Sätze Einfache Sätze

Geschäftlich Privat

Gespräch mit Arbeitskollegen Kundengespräch

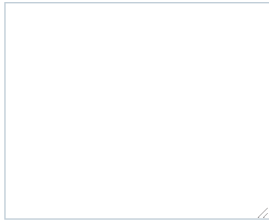
21 Geschrieben wie Ideal

Bei der letzten Frage haben Sie beschrieben, wie man bei dieser Aufgabe schreiben sollte. Haben Sie immer genau so geschrieben?

- Ja
- Nein
-

22.1 Anders geschrieben als Ideal

Aus welchen Gründen haben Sie anders geschrieben?



23 Textart Gesprächspartner

Wie hat Ihr Gesprächspartner geschrieben?

Bitte wählen Sie in jeder Zeile die Beschreibung aus, die besser auf die Schreibart Ihres Gesprächspartners zutrifft.

Mündlicher Stil Schriftlicher Stil

Auf Höflichkeitsformen verzichten Höflichkeitsformen anwenden

Abwechslungsreiche Wörter und Satzstellungen Möglichst einfach verständliche Wörter und Satzstellungen

Business English Alltagsenglisch

Gespräch mit Arbeitskollegen Kundengespräch

Komplexe Sätze Einfache Sätze

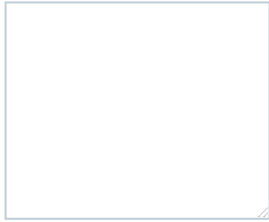
Geschäftlich Privat

Auf Rechtschreibung achten Nicht auf Rechtschreibung achten

Formell Informell

24 Abschlussfrage

Haben Sie Fragen oder Bemerkungen?

An empty rectangular box with a thin black border, intended for the user to write questions or remarks. The box is currently blank.

25 Endseite

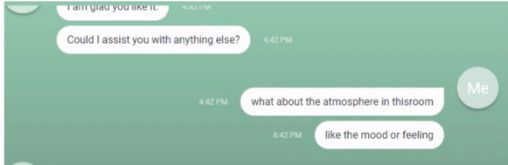
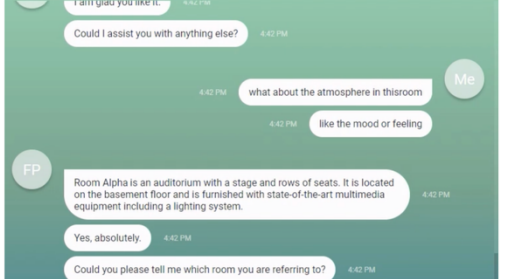
Herzlichen Dank für Ihre Teilnahme!

Appendix E Example interview guide

KM8-5TV

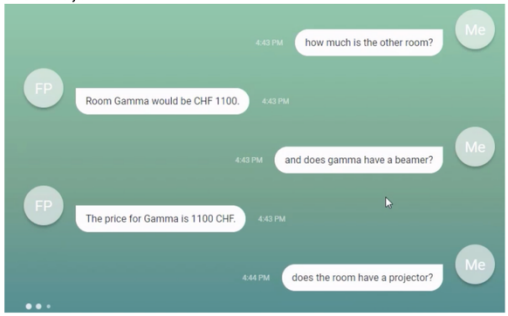
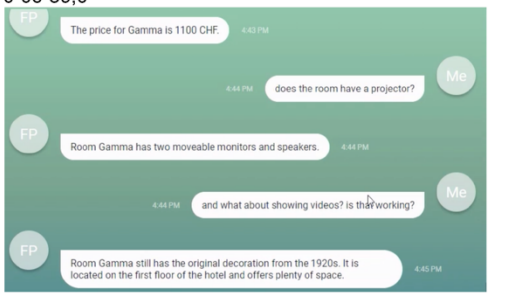
10.5.2019
Dauer ... Min
WKS, ...

Interview Protocol KM8-5TV

Stimulated Recall		
Episode	Leading and focus question	Retrospection
<p>0-04-00,0</p> 	<p><i>Was waren Ihre Überlegungen an dieser Stelle?</i></p> <p><i>Was überlegten Sie sich, als Sie das schrieben?</i></p>	
<p>0-04-12,0</p> 	<p><i>Was dachten Sie in diesem Moment?</i></p> <p><i>Was überlegten Sie sich, als der Chatbot das schrieb?</i></p>	

KM8-STV

10.5.2019
Dauer ... Min
WKS, ...

<p>0-05-09,6</p> 	<p><i>Was waren Ihre Überlegungen hier?</i></p> <p><i>Was dachten Sie, als Sie schrieben, dann nach oben scrollten und weiterschrieben?</i></p>	
<p>0-06-39,0</p> 	<p><i>Was waren Ihre Überlegungen hier?</i></p> <p><i>Was dachten Sie, als der Chatbot das schrieb?</i></p>	

KM8-STV

10.5.2019
Dauer ... Min
WKS, ...

<p>0-07-01,0</p> 	<p><i>Was waren Ihre Überlegungen in diesem Moment?</i></p> <p><i>Was dachten Sie, als Sie das schrieben?</i></p>	
<p>0-07-49,0</p> 	<p><i>Was dachten Sie an dieser Stelle?</i></p> <p><i>Was dachten Sie, als Sie das schrieben?</i></p>	

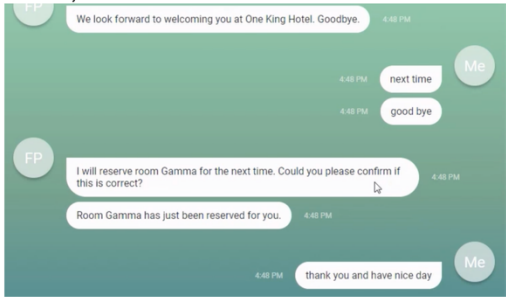
KM8-STV

10.5.2019
Dauer ... Min
WKS, ...

<p>0-08-19,0</p> 	<p><i>Was waren Ihre Überlegungen in diesem Moment?</i></p> <p><i>Was überlegten Sie sich, als Sie das schrieben?</i></p>	
<p>0-09-10,0</p> 	<p><i>Was dachten Sie in diesem Moment?</i></p> <p><i>Was dachten Sie, als Sie das schrieben?</i></p>	

KM8-STV

10.5.2019
 Dauer ... Min
 WKS, ...

<p>0-09-30,0</p> 	<p><i>Was dachten Sie an dieser Stelle?</i></p> <p><i>Was waren Ihre Überlegungen als der Chatbot das schrieb?</i></p> <p><i>Was dachten Sie, als Sie das schrieben?</i></p>	
---	--	--

KM8-STV

10.5.2019
Dauer ... Min
WKS, ...

Semistructured Interview	
Question	Answer
<i>Was dachten Sie über die Aufgabe?</i>	
<i>Was fiel Ihnen auf beim Ausführen des Auftrags, also beim Auswählen eines passenden Raums? Können Sie mir noch mehr dazu erzählen, wie Sie vorgegangen sind?</i>	
<i>Was fiel Ihnen am Chatten auf? Was war anders, als wenn Sie sonst chatten?</i>	
<i>Was fiel Ihnen an Ihrem Gesprächspartner auf? Was fiel Ihnen z.B. am Schreibstil, am Englisch, an der Kommunikation auf?</i>	
<i>Was fiel Ihnen an sich selbst auf? Was fiel Ihnen z.B. an Ihrem Schreibstil, Englisch, an der Kommunikation auf?</i>	
<i>Haben wir etwas vergessen, was Sie gerne noch ansprechen würden? Gibt es noch etwas, das Sie gerne ansprechen</i>	

Appendix F Test administration scripts

Week 1

Testleiterskript Woche 1

Mein Name ist Johanna Oeschger. Ich bin Berufsfachschullehrerin für Englisch und doktoriere am Institut für Bildungswissenschaften der Universität Basel und Fachhochschule Nordwestschweiz.

Im Rahmen meines Doktorats führe ich ein Forschungsprojekt durch mit dem Titel «Chatbots in the English Classroom». In dem Forschungsprojekt möchten wir prüfen, ob man Englisch-Aufgaben auch mit modernen Kommunikationsformen, nämlich über Chat, durchführen kann.

Vielen Dank, dass ich im Rahmen dieses Forschungsprojekts eine Studie in Ihrer Klasse durchführen darf.

Die Studie wird in drei Teilen durchgeführt:

- Heute stelle ich Ihnen die Studie vor und Sie absolvieren einen kurzen Englisch Sprachtest.
- Nächste Woche werden Sie die Aufgabe ausführen. Vor und nach der Aufgabe werden Sie jeweils einen kurzen Fragebogen zur Aufgabe ausfüllen.
- Übernächste Woche möchte ich mit 3-4 Personen Ihrer Klasse ein Interview durchführen. Im Interview werde ich Ihnen Fragen stellen zu Ihren Erfahrungen mit der Aufgabe.

Ihre Teilnahme an dieser Studie ist freiwillig. Die Daten werden anonymisiert erhoben, gespeichert und ausgewertet. Das heisst, Ihre Angaben werden mit einem Code gekennzeichnet und können bei der Auswertung und Publikation nicht mit Ihrem Namen in Verbindung gebracht werden.

Für den Erfolg unseres Projekts ist Ihre Teilnahme an dieser Studie sehr wichtig und Sie können mit Ihrer Teilnahme für die Weiterentwicklung des Englischunterrichts an der Berufsfachschule beitragen.

Wenn Sie mit der Teilnahme an der Studie einverstanden sind, bitte ich Sie, die Einverständniserklärung zu unterzeichnen.

Sprachtest

- Der Sprachtest hilft uns, Ihre momentanen Englischkenntnisse einzuschätzen und so die Ergebnisse der Studie besser auswerten zu können.
- Das Testergebnis wird ausschliesslich für die Studie verwendet und hat keinen Einfluss auf Ihre Englischnote.
- Es ist wichtig, dass der Sprachtest ein möglichst genaues Bild Ihrer aktuellen Englischkompetenzen gibt. Bitte lösen Sie deshalb den Test so gut Sie können, ohne fremde Hilfe und ohne Hilfsmittel.
- Sie haben für den Sprachtest 30 Minuten Zeit. Es ist nicht zwingend, dass Sie nach Ablauf der Zeit fertig sind mit der Aufgabe. Falls Sie früher fertig sind, bearbeiten Sie bitte Teil 2 des Sprachtests.

Week 2

Testleiterskript Woche 2

Beim heutigen Teil der Studie geht es darum, eine neue Aufgabe für den Englischunterricht in der Praxis zu testen.

Ablauf

- Sie werden zuerst einen ersten Fragebogen ausfüllen.
- Anschliessend führen Sie die Aufgabe aus.
- Nachdem Sie die Aufgabe abgeschlossen haben, füllen Sie einen zweiten Fragebogen aus.

Nach jedem Schritt gibt es eine kurze Pause und ich gebe Ihnen jeweils eine Anleitung zum nächsten Schritt.

Pre-Fragebogen

- Bitte klicken Sie im Browser auf das Lesezeichen «Umfrage 1»
- Geben Sie auf der ersten Seite des Fragebogens bitte Ihren persönlichen Code ein im Format XXX-XXX. Der Code steht auf Ihrem Auftragsblatt.
- Das Ausfüllen des Fragebogens dauert etwa 5 Minuten. Bleiben Sie nach Ausfüllen des Fragebogens bitte an Ihrem Platz.

Aufgabe

- Bei der Aufgabe handelt es sich um ein Rollenspiel.
- Sie führen das Rollenspiel über Chat durch. Es geht darum zu prüfen, ob man die Aufgaben auch mit modernen Kommunikationsformen durchführen kann.
- Eine Gruppe von Ihnen chattet mit einer Kollegin bzw. einem Kollegen aus der Klasse. Die andere Gruppe chattet mit einem Chatbot, der speziell für die Aufgabe entwickelt wurde. Die Gruppenzuteilung erfolgt zufällig.
- Bitte schreiben Sie beim Chatten ausschliesslich auf Englisch und benutzen Sie keine Hilfsmittel (Wörterbuch, Übersetzungsdienst, Google, Sitznachbar/in, ...). Verwenden Sie beim Chatten keine Emojis.
- Sie haben für die Aufgabe 20 Minuten Zeit. Es ist nicht zwingend, dass Sie nach Ablauf der Zeit fertig sind mit der Aufgabe. Falls Sie früher fertig sind, bearbeiten Sie bitte das Arbeitsblatt an Ihrem Platz.
- Die Protokolle Ihres Chats sowie eine Videoaufnahme Ihres Bildschirms werden gespeichert und anschliessend ausgewertet.
- Bitte lesen Sie die Aufgabe jetzt durch und fragen Sie mich, wenn etwas unklar ist.
- Klicken Sie im Internetbrowser auf das Lesezeichen «Chat» resp. «Chatbot» und beginnen Sie mit der Aufgabe.
- Um den Chat mit dem Chatbot zu beginnen, geben Sie bitte zuerst Ihren persönlichen Code ein.

Post-Fragebogen

- Bitte klicken Sie im Internetbrowser auf das Lesezeichen «Umfrage 2».
- Geben Sie auf der ersten Seite des Fragebogens bitte Ihren persönlichen Code ein im Format XXX-XXX. Der Code steht auf Ihrem Laptop.
- Das Ausfüllen des Fragebogens dauert etwa 10 Minuten. Bleiben Sie nach Ausfüllen des Fragebogens bitte an Ihrem Platz.

Week 3

Testleiterskript Woche 3

Engage in some small talk, then start the audio recording.

Present the interview procedure:

Sie haben letzte Woche eine neue Aufgabe für den Englischunterricht ausprobiert. Ich möchte Ihnen heute Fragen stellen dazu.

Es ist wichtig, dass Sie so antworten, wie Sie wirklich denken und fühlen. Es gibt keine richtigen oder falschen Antworten und auch keine Antworten, die mir lieber sind als andere. Was Sie heute sagen, wirkt sich nicht auf Ihre Englischnote oder sonst in irgendeiner Weise auf Ihre schulische Bewertung aus.

Wir schauen uns nun gemeinsam Ausschnitte aus Ihrem Chat von letzter Woche an. Mich interessiert, was Ihre Überlegungen und Gedanken beim Chatten waren. Ich kann im Chat-Protokoll und auf den Bildschirmaufnahmen sehen, was Sie während der Aufgabe geschrieben und getan haben, aber ich weiss nicht, was Sie dabei dachten.

Ich möchte deshalb heute von Ihnen wissen, was Ihre Gedanken und Überlegungen beim Chatten letzte Woche waren. Versuchen Sie, sich zu erinnern, was Sie letzte Woche beim Chatten dachten, nicht, was Sie jetzt darüber denken.

Ich zeige Ihnen jetzt einige Ausschnitte aus den Bildschirmaufnahmen. Anschliessend werde ich Sie bitten, sich zurück zu erinnern und mir zu sagen, was Sie in dem Moment dachten.

Give the participant choice of language:

Möchten Sie das Interview lieber auf Standarddeutsch oder Schweizerdeutsch weiterführen?

Stimulated Recall

Present one video segment at a time. Ask a general question about it, for example:

*Was dachten Sie an dieser Stelle / hier / in diesem Moment?
Was waren Ihre Überlegungen in diesem Moment?*

If the participant responds before the video segment has finished playing, pause the video. After the participant has completed his/her response, play the rest of the segment and ask:

*Erinnern Sie sich sonst noch an etwas, das Sie an dieser Stelle dachten?
Was dachten Sie sonst noch an dieser Stelle?*

If the participant's response is that they don't remember, do not 'fish' for answers.

If uncertain whether the participant comments on time of production, ask for example:

Dachten Sie das letzte Woche während dem Chatten oder denken Sie das heute?

If needed, ask open follow-up questions for more detailed description or clarification, for example:

*Und dann?
Können Sie mehr dazu sagen?
Können Sie das noch näher/ausführlicher beschreiben?
Das habe ich nicht so ganz verstanden. Was waren Ihre Überlegungen da?*

In follow-up questions, focus attention on specific parts or aspects of the segment related to the research question (task, genre (structure, syntax, lexicon, pragmatics), miscommunication/negotiations, self-correction/repair etc.), for example:

*Mir fällt auf, dass Sie ... Was dachten Sie dabei?
Was überlegten Sie sich, als Sie löschten und neu schrieben?
Was dachten Sie, als der Chatbot/Gesprächspartner das sagte?
Was waren Ihre Überlegungen, als Sie das schrieben?*

After the response, wait for a few seconds, then ask:

An welche Gedanken oder Überlegungen erinnern Sie sich sonst noch?

React to responses by simple, encouraging backchanneling cues or nonresponses:

Ok, ah, mhm
Nodding, eye contact

Semi-structured interview

After simulated recall is complete, move to semi-structured interview. Ask all questions, even if the participant already provided answers during simulated recall:

1. *Was dachten Sie über die Aufgabe?*
2. *Was fiel Ihnen auf beim Ausführen des Auftrags, also beim Auswählen eines passenden Raums?*
3. *Was fiel Ihnen am Chatten auf?*
Follow-up question: *Was war anders, als wenn Sie sonst chatten?*
4. *Was fiel Ihnen an Ihrem Gesprächspartner auf?*
Follow-up question: *Was fiel Ihnen z.B. am Schreibstil, am Englisch, an der Kommunikation auf?*
5. *Was fiel Ihnen an sich selbst auf?*
Follow-up question: *Was fiel Ihnen z.B. an Ihrem Schreibstil, Englisch, an der Kommunikation auf?*

If needed, ask open follow-up questions for more detailed description or clarification, for example:

*Können Sie mehr dazu sagen?
Können Sie das noch näher/ausführlicher beschreiben?
Können Sie ein Beispiel für ... nennen?
Das habe ich nicht so ganz verstanden. Was waren Ihre Überlegungen da?*

After the response to each question, wait for a few seconds, then ask:

*Was dachten Sie sonst noch?
Was fiel Ihnen sonst noch auf?*

Maybe ask additional questions related to the research question:

- Genre: social roles, structure, style, style change, pragmatics, ...
- Language use: noticing, noticing the gap, modified speech, self-corrections, negotiations of meaning, perceived language skills, ...
- Communication: misunderstandings, strategies, non-understanding by chatbot, ...

Sie sagten, der Chatbot habe Sie oft nicht verstanden. Können Sie mehr dazu sagen? Wie fühlten Sie sich in diesen Momenten?

Sie haben während des Interviews ein paar Mal Fehler angesprochen, die Sie beim Chatten machten. Können Sie das noch näher beschreiben?

Sie erinnerten sich im Interview, dass Ihnen während des Chattens mehrmals das richtige Wort nicht in den Sinn gekommen war. Können Sie mehr dazu sagen?

After semi-structured interview is complete, ask concluding question, for example:

*Haben wir etwas vergessen, was Sie gerne noch ansprechen würden?
Gibt es noch etwas, das Sie gerne ansprechen würden?*

Thank participant for their time and turn off audio recording.

Appendix G Student consent form

Johanna Oeschger
johanna.oeschger@unibas.ch

Institut für Bildungswissenschaften
Universität Basel

Einverständniserklärung Projekt «Chatbots in the English Classroom»

Ich, _____, (Name und Vorname)
bestätige, dass ich mit der Erhebung und Verwendung der von mir erhobenen Daten im
Projekt «Chatbots in the English Classroom» einverstanden bin.

- Ich wurde über den Ablauf und Zweck des Projekts informiert und erkläre hiermit meine freiwillige Teilnahme an diesem Forschungsprojekt.
- Ich bin damit einverstanden, dass meine Daten in anonymisierter Form aufgezeichnet, gespeichert und ausgewertet werden.
- Ich bin ebenso damit einverstanden, dass Ausschnitte meiner Aussagen in Publikationen im Rahmen des Projekts veröffentlicht werden. Jede Zitierung erfolgt vollständig anonymisiert und dient ausschliesslich wissenschaftlichen Zwecken.

Ort, Datum

Unterschrift

Appendix H Codebooks derived from interaction scripts

Both codebooks (Codebook for interactional features and Codebook for complexity, accuracy, communicative adequacy) may be accessed at Zenodo repository:

<https://doi.org/10.5281/zenodo.13341676>

Appendix I Complete intercoder agreement measures

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
INTERACTIONAL FEATURES						
Recasts						
Recast move		8	5	13	61.54	-
Linguistic focus	Spelling	6	0	6	100.00	1.00
	Lexis	2	0	2	100.00	1.00
	Grammar	8	0	8	100.00	1.00
	Multi-focus	-	-	-	-	-
	Total	16	0	16	100.00	1.00
Intensity of recast	Singular incidence	12	0	12	100.00	1.00
	Repeated incidence	4	0	4	100.00	1.00
	Total	16	0	16	100.00	1.00
Intensity of recast form	Singular incidence	4	0	4	100.00	1.00
	Repeated incidence	12	0	12	100.00	1.00
	Total	16	0	16	100.00	1.00
Completeness	Complete	6	0	6	100.00	1.00
	Partial	10	0	10	100.00	1.00
	Total	16	0	16	100.00	1.00
Length	Single sentence	6	0	6	100.00	1.00
	Multiple sentences	4	0	4	100.00	1.00
	Multiple messages	6	0	6	100.00	1.00
	Total	16	0	16	100.00	1.00
Uptake	No uptake	12	0	12	100.00	1.00
	Immediate uptake\Repair	-	-	-	-	-
	Immediate uptake\Needs repair	-	-	-	-	-
	Delayed uptake\Repair	-	-	-	-	-

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
	Delayed uptake\Repair	2	1	3	66.67	0.61
	'Re-uptake'\Repair	0	1	1	0.00	0.00
	'Re-uptake'\Repair	0	1	1	0.00	0.00
	Total	14	3	17	82.35	0.79
Negotiation of meaning						
Trigger	Erroneous	34	1	35	97.14	0.94
	Target-like	30	1	31	96.77	0.94
	Total	64	2	66	96.97	0.94
Indicator		48	11	59	81.36	-
Initiator	Learner	2	0	2	100.00	1.00
	Interlocutor	64	0	64	100.00	1.00
	Total	66	0	66	100.00	1.00
Type	Clarification request	20	1	21	95	0.94
	Confirmation check	4	1	5	80.00	0.73
	Inappropriate response	34	2	36	94.44	0.93
	Overlap	4	0	4	100.00	1.00
	Total	62	4	66	93.94	0.92
Response						
Type	Meta-comment	-	-	-	-	-
	Change topic	12	1	13	92.31	0.92
	Accept topic	14	1	15	93.33	0.93
	Elaboration	16	2	18	88.89	0.88
	Reduction	10	0	10	100.00	1.00
	Modification	6	2	8	75.00	0.73
	Repetition	2	0	2	100.00	1.00
	No response	-	-	-	-	-
	Total	60	6	66	90.91	0.90

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
Accuracy of reformulation	Improved	14	0	14	100.00	1.00
	Reduced	8	0	8	100.00	1.00
	Equivalent	14	0	14	100.00	1.00
	Total	36	0	36	100.00	1.00
Reaction to response	Implicit resolution	42	2	44	95.45	0.94
	Implicit no resolution	8	2	10	80.00	0.75
	Explicit no resolution	8	3	11	72.73	0.66
	No reaction to response	-	-	-	-	-
	Total	58	7	65	89.23	0.87
Complexity	Single-layered	24	0	24	100.00	1.00
	Multi-layered	18	0	18	100.00	1.00
	Total	42	0	42	100.00	1.00
Structural alignment						
Copula <i>be</i>		332	16	348	95.40	-
Direction of alignment	Learner > learner	4	0	4	100.00	1.00
	Learner > interlocutor	24	1	25	96.00	0.95
	Learner > both	88	2	90	97.78	0.97
	Learner unprimed	8	0	8	100.00	1.00
	Interlocutor > learner	138	0	138	100.00	1.00
	Interlocutor > interlocutor	76	0	76	100.00	1.00
	Interlocutor unprimed	14	0	14	100.00	1.00
	Total	228	0	228	100.00	1.00
<i>Have</i> + noun phrase		110	8	118	93.22	-
Direction of alignment	Learner > learner	4	0	4	100.00	1.00
	Learner > interlocutor	12	0	12	100.00	1.00
	Learner > both	14	0	14	100.00	1.00
	Learner unprimed	10	0	10	100.00	1.00

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
	Interlocutor > learner	40	0	40	100.00	1.00
	Interlocutor > interlocutor	16	1	17	94.12	0.93
	Interlocutor unprimed	20	1	21	95.24	0.94
	Total	76	2	78	97.44	0.97
<i>For + noun phrase</i>		214	2	216	99.07	-
Direction of alignment	Learner > learner	14	0	14	100.00	1.00
	Learner > interlocutor	20	0	20	100.00	1.00
	Learner > both	30	1	31	96.77	0.96
	Learner unprimed	10	0	10	100.00	1.00
	Interlocutor > learner	80	1	81	98.77	0.99
	Interlocutor > interlocutor	28	2	30	93.33	0.92
	Interlocutor unprimed	20	1	21	95.24	0.94
	Total	128	4	132	96.97	0.96
Modal + infinitive		242	6	248	97.58	-
Direction of alignment	Learner > learner	8	2	10	80.00	0.77
	Learner > interlocutor	28	0	28	100.00	1.00
	Learner > both	40	1	41	97.56	0.97
	Learner unprimed	-	-	-	-	-
	Interlocutor > learner	58	1	59	98.31	0.98
	Interlocutor > interlocutor	76	0	76	100.00	1.00
	Interlocutor unprimed	22	0	22	100.00	1.00
	Total	156	1	157	99.36	0.99
<i>To-infinitive</i>		94	0	94	100.00	-
Direction of alignment	Learner > learner	10	1	11	90.91	0.89
	Learner > interlocutor	6	0	6	100.00	1.00
	Learner > both	18	1	19	94.74	0.94
	Learner unprimed	12	0	12	100.00	1.00

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
	Interlocutor > learner	16	0	16	100.00	1.00
	Interlocutor > interlocutor	4	0	4	100.00	1.00
	Interlocutor unprimed	12	0	12	100.00	1.00
	Total	32	0	32	100.00	1.00
Existential <i>there</i>		16	0	16	100.00	-
Direction of alignment	Learner > learner	-	-	-	-	-
	Learner > interlocutor	-	-	-	-	-
	Learner > both	-	-	-	-	-
	Learner unprimed	-	-	-	-	-
	Interlocutor > learner	-	-	-	-	-
	Interlocutor > interlocutor	-	-	-	-	-
	Interlocutor unprimed	-	-	-	-	-
	Total	-	-	-	-	-
Question with subject-auxiliary inversion		104	11	115	90.43	-
Direction of alignment	Learner > learner	8	0	8	100.00	1.00
	Learner > interlocutor	16	0	16	100.00	1.00
	Learner > both	10	0	10	100.00	1.00
	Learner unprimed	8	0	8	100.00	1.00
	Interlocutor > learner	14	0	14	100.00	1.00
	Interlocutor > interlocutor	32	0	32	100.00	1.00
	Interlocutor unprimed	24	0	24	100.00	1.00
	Total	70	0	70	100.00	1.00
<i>Wh</i> -interrogative		84	6	90	93.33	-
Direction of alignment	Learner > learner	24	0	24	100.00	1.00
	Learner > interlocutor	4	0	4	100.00	1.00
	Learner > both	2	0	2	100.00	1.00
	Learner unprimed	10	1	11	90.91	0.89
	Interlocutor > learner	6	0	6	100.00	1.00

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
	Interlocutor > interlocutor	12	0	12	100.00	1.00
	Interlocutor unprimed	18	1	19	94.74	0.94
	Total	36	1	37	97.30	0.97
Lexical alignment						
Direction of alignment	Learner > learner	56	4	60	93.33	0.92
	Learner > interlocutor	16	1	17	94.12	0.93
	Learner > both	10	0	10	100.00	1.00
	Learner unprimed	136	6	142	95.77	0.95
	Interlocutor > learner/With error	20	0	20	100.00	1.00
	Interlocutor > learner/Error-free	36	1	37	97.30	0.97
	Interlocutor unprimed	48	2	50	96.00	0.95
	Total	104	3	107	97.20	0.97
COMPLEXITY AND ACCURACY						
Errors	Grammar	80	27	107	74.77	-
	Spelling	70	10	80	87.50	-
	Lexis	64	21	85	75.29	-
	Total	214	58	272	78.68	-
AS-units	Error-free	238	30	268	88.81	-
	With error	146	21	167	87.43	-
	Total	384	51	435	88.28	-
Requests		178	33	211	84.36	-
Request strategy	Imperative	2	0	2	100.00	1.00
	Direct question	76	14	90	84.44	0.82
	Elliptical question	12	3	15	80.00	0.77
	Indirect question	2	1	3	66.67	0.62
	Want statement	14	2	16	87.50	0.86
	Assertion	8	1	9	88.89	0.87

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
	Reference to preparatory condition	6	4	10	60.00	0.54
	Strong hint	60	5	65	92.31	0.91
	Total	180	30	210	85.71	0.84
Internal modifier	Politeness marker	2	0	2	100.00	1.00
	Modal preterite	36	3	39	92.31	0.85
	Total	38	3	41	92.68	0.85
External modifier	Alerter	6	1	7	85.71	0.82
	Grounder	8	2	10	80.00	0.75
	Agreer	14	1	15	93.33	0.92
	Appreciator	4	0	4	100.00	1.00
	Disagreer	2	0	2	100.00	1.00
	Total	34	4	38	89.47	0.87
Appropriateness	Inappropriate	32	9	41	78.05	0.67
	Partly appropriate	108	13	121	89.26	0.84
	Fully appropriate	44	4	48	91.67	0.88
	Total	184	26	210	87.62	0.81
COMMUNICATIVE ADEQUACY						
Task completion	Successful	14	0	14	100.00	1.00
	Partly successful	2	0	2	100.00	1.00
	Not successful	-	-	-	-	-
	Total	16	0	16	100.00	1.00
Generic structure	Fully genre-appropriate	2	1	3	66.67	0.50
	Partly genre-appropriate	12	1	13	92.31	0.88
	Inappropriate	-	-	-	-	-
	Total	14	2	16	87.50	0.81
Generic style	Formal	4	0	4	100.00	1.00
	Mixed	8	1	9	88.89	0.83

Main code	Subcode	Agreements	Disagreements	Total	Percent agreement	Kappa (Brennan & Prediger, 1981)
	Informal	2	1	3	66.67	0.50
	Total	14	2	16	87.50	0.81

Appendix J Grouped responses to questions in post-questionnaire

Table J.1

Frequency and Distribution of Grouped Responses to the Question “What Did You Do When You Couldn’t Understand Your Conversation Partner?”

Grouped response	System (n = 21)		SCMC (n = 18)	
	Count	%	Count	%
Derived meaning from context	3	14.3	3	16.7
Checked in chat	2	9.5	1	5.6
Changed topic	2	9.5	0	0.0
Asked interlocutor to explain	1	4.8	1	5.6
Looked up online	0	0.0	1	5.6
Other	0	0.0	1	5.6
Told interlocutor that not understood	0	0.0	0	0.0
Asked interlocutor if understood correctly	0	0.0	0	0.0
Did nothing	0	0.0	0	0.0

Table J.2

Frequency and Distribution of Grouped Responses to the Question “What Did You Do When Your Conversational Partner Couldn’t Understand You?”

Grouped response	System (n = 21)		SCMC (n = 18)	
	Count	%	Count	%
Used a different word	9	42.9	3	16.7
Wrote more detailed	7	33.3	1	5.6
Changed spelling	6	28.6	1	5.6
Changed syntax	6	28.6	0	0.0
Repeated message	5	23.8	0	0.0
Split statement into multiple messages	4	19.0	0	0.0
Other	4	19.0	0	0.0
Changed topic	3	14.3	0	0.0
Wrote less detailed	2	9.5	0	0.0

Grouped response	System (n = 21)		SCMC (n = 18)	
	Count	%	Count	%
Summarized multiple messages into one	2	9.5	0	0.0
Asked interlocutor if understood	1	4.8	0	0.0
Wrote in German	0	0.0	0	0.0

Appendix K Codes pertaining to negotiation of meaning derived from learner interviews

Main code	Subcode	Definition	Example
Emotional reaction	Frustrated/Annoyed	The learner was annoyed or frustrated about the interlocutor's lack of understanding.	[...] also, wenn er mich nicht verstand, war ich schon etwas genervt, [...]
	Lost patience	The learner lost patience to continue the conversation because of the interlocutor's non-understanding.	da löschte es mir wohl etwas ab, eben weil ich [...] das Problem hatte mit dem Roboter
	Confused	The learner was confused as they could not make sense of the interlocutor's inadequate response.	Ich verstand dann auch nicht ganz, [...] als ich ihm sagte, welchen Raum ich will, dass er mir noch einmal sagte, dass der auch, ähm, Bildschirme hat und Lautsprecher, weil das hatte er mir ja eigentlich schon vorher gesagt.
	Curious	The learner was curious to test out the limits of the interlocutor.	neugierig. (...) Also, was er, was jetzt die Antwort sein wird.
	Amused	The learner was amused about the interlocutor's lack of understanding.	in dem Moment dachte ich, ah, was für ein dummer Computer. (..) Und fand es lustig ((lacht))
Learner conceptions	Interlocutor's fault		
	Long message	The dialog system cannot process longer messages or messages containing multiple requests.	wenn man [...] mal eine längere Frage hat oder mehrere Fragen in einer Nachricht, dann antwortet er einfach auf eine Frage von den drei oder vier.
	First time	The dialog system sometimes does not understand the first time.	Und dann fragte ich noch einmal [...] in der Hoffnung, dass er diesmal antwortet.
	Retaining of information	The dialog system is not capable of retaining information, i.e., does not have a 'memory'.	das Problem ist, es kann nicht speichern, was man vorher wie gesagt hatte [...] dann musst du einfach immer wiederholen und ein paar Dinge dann nicht, aber ein paar Dinge dann schon

Main code	Subcode	Definition	Example
	Vocabulary	The dialog system cannot understand all words, i.e., has a limited vocabulary.	Also ja, das Wort "amazing" verstand er wahrscheinlich nicht, nehme ich an.
	Keyword-based	The dialog system is purely keyword-based, i.e., it does not consider a message's context beyond keywords to establish its meaning.	das Programm ist vielleicht nicht ganz so (..) äh, (.) gut programmiert [...] also, es hat wahrscheinlich einfach so (.) auf einzelne Fragen, (.) [...] wo vermutlich paar Stichwörter kommen, so paar Antworten programmiert, die dann gleich kommen, (.) und nicht [...] auf den ganzen Text bezogen.
	Learner's fault		
	Unspecified	The learner does not specify what aspect of their production might have led to non-understanding.	aber wahrscheinlich formulierte ich es irgendwie falsch, so wie es dann verstanden wurde [...]
	Inadequate discourse management	Non-understanding due to inadequate discourse management.	Also, jetzt von meiner Seite aus, machte ich es, glaube ich, (...) überkompliziert, also ich schrieb einfach zurück. Ich hätte schneller zum Punkt kommen können.
	Lexical error	Non-understanding due to a lexical error such as lexical misselection or use of German.	Oder auch-- weil es einfach das falsche Wort war.
	Unclear reference	Non-understanding due to unclear reference.	Ja, dass ich eigentlich lieber hätte wissen wollen, wie teuer die Räume waren und dass ich die Frage eigentlich auf diejenige von oben bezogen hatte, aber er hat das dann wohl erst auf den Rauminhalt sozusagen (.) verknüpft, und nicht auf den Raumpreis. Ich hätte vielleicht auch etwas, äh, genauer nach dem Preis fragen sollen.

Main code	Subcode	Definition	Example
Learner strategies	Restart	The learner restated an introductory utterance to 'restart' the conversation.	und dann habe ich einfach noch einmal wie von Anfang angefangen, was überhaupt der Grund ist, warum ich ihm schreibe. Also, ich habe dann noch einmal geschrieben, dass ich jetzt gerne noch den Raum reservieren und buchen würde
	Repeat	The learner repeated the message that created non-understanding without modifying it.	Und dann fragte ich noch einmal einfach
	Meta-comment	The learner addressed the interlocutor on a metalinguistic level, i.e., asking the interlocutor to specify or to answer a question.	Dann fragte ich einfach einmal, ob er es irgendwie per Mail oder wie er es will.
	Keyword	The learner used a single keyword to express the central meaning of the message that created non-understanding.	[...] wahrscheinlich "foot" und "food" falsch verstanden hatte, schrieb ich dann noch einmal "food"
	All-in-one	The learner added multiple pieces of information into a single message.	Ich dachte mir dann, ich probiere vielleicht gerade alle Informationen auf einmal zu geben,
	Step-by-step	The learner composed messages that contained only single pieces of information instead of multiple pieces of information.	ich habe einfach immer nur eine Antwort nach dem anderen, weil ich mir dachte, er versteht es nicht mehr, weil ich zu viel schreibe oder so.
	Rephrase	The learner rephrased the message that created non-understanding. The rephrasal may additionally involve elaboration or reduction of the information content of the trigger message.	und dann versuchte ich es, so, noch ein bisschen vereinfacht noch einmal zu senden
	Give up	The learner made no attempt to resolve the non-understanding and changed the topic or left the interaction.	das verstand er nicht. Da dachte ich, egal ((lacht)).

Main code	Subcode	Definition	Example
			Und das zweite Mal liess ich es dann eigentlich gleich bleiben und stellte eine andere Frage.

Note. All subcodes solely reflect the perspective of learners from the learner-system group on the learner-system interaction type, as learner responses on negotiation of meaning were absent in the SCMC interviews. With regard to the interview findings, the code designations and definitions are presented in the original English while the code examples are presented in the original – or transcribed from Swiss German, respectively – Standard German. The interview extracts presented as code examples were also shortened for better readability, with ellipses indicated by three periods in square brackets, i.e., [...].

Appendix L Codes pertaining to complexity derived from learner interviews

Main code	Subcode	Definition	Example	
			System	SCMC
Focus	Message-level	Focus on complexity at the level of an entire message such as the number of requests per message.	ich habe einfach immer nur eine Antwort nach dem anderen, weil ich mir dachte, er versteht es nicht mehr, weil ich zu viel schreibe oder so. Also schrieb ich dann wirklich ganz einfach	Und dann halt mehrere untereinander statt in einem schreiben, was man gerne hätte, damit die andere Person nicht so viele (...) Sätze, Wörter hat.
	Sentence-level	Focus on complexity at the sentence-level such as sentence length or the general complexity of a sentence.	Ich probierte [...] jetzt nicht irgendwelche komplizierten Sätze zu schreiben	Ja, und wir schrieben in einfachen Sätzen, so war es noch einfach, es zu verstehen
	Word-level	Focus on complexity at the word-level, i.e., relating to lexical diversity.	-	ich versuchte schon, mir-- mir Mühe zu geben, aber wenn man halt im Chat ist, hat man so, wie einen kleinen gewissen Zeitdruck und da überlegt man sich nicht (.) noch weitere Wörter oder irgendwie Synonyme für ein Wort (.) und darum, denke ich, machen wir es alle so einfach wie möglich.

Main code	Subcode	Definition	Example	
			System	SCMC
	Not specified	The learner does not specify the aspect on which their statement about complexity focuses.	Und eigentlich wechselte ich nie viel, also schrieb ich es einfach immer wie einfacher. Ich gab mir immer weniger um-- Mühe.	[...] versuchte immer, so ein bisschen einfach alles zu formulieren, damit es für die andere Person möglichst einfach ist und für mich auch möglich einfach, damit die andere Person dann auch nicht mega formell oder so schreibt
Reasons for enhancing complexity	Email as model	To model the production after an email text.	[...] dann entschied ich mich eben, so wie ein Mail-- also erst einmal, (..) ähm, viel Text (..) und nicht-- äh-- und dann gleich alle Informationen so	-
	Increase information content	To increase the information content of the message, e.g., to ensure that the dialog system has all the information.	ich wollte ja immer etwas Kleines und dann vergass es wie das Vorherige und ich war dann nicht sicher, hat er das jetzt gespeichert, also-- werde ich jetzt den Punkt bekommen, weil ich das jetzt gelöst habe oder nicht, also wie-- wenn ich mir vorstelle, dass er das jetzt so re-- äh, reserviert, wird er das jetzt richtig reservieren. Dann dachte	-

Main code	Subcode	Definition	Example	
			System	SCMC
	Apply Business English	To apply Business English in order to enhance learning.	-	<p>ich mir, ich mache jetzt einmal einen ganzen Text</p> <p>Aber ich merkte, dass er sich Mühe gibt und, ähm, die Sätze einfach-- so einfach wie möglich fasste. Was, ähm, ja schon praktisch ist, also man versteht es dann besser, kommt besser voran, aber (..) man entwickelt sich, glaube ich, nicht so. Also, man macht es so-- so einfach, so basic wie möglich und versucht, das Business gar nicht anzuwenden.</p>
	Not specified	The learner does not specify the reason for enhancing the complexity of their production.	-	<p>also bei diesem Auftrag machte ich natürlich in längeren-- längere Sätze gemacht. Als privat sind wir-- fassen wir uns eigentlich recht kurz beim Chatten.</p>
Reasons for reducing complexity	Lack of competence	Lacked the competence to produce complex language.	dann wusste ich nicht, wie man es eigentlich richtig schreibt. Also vereinfachte ich es dann viel.	wir schrieben in einfachen Sätzen, so war es noch einfach, es zu verstehen und auch die Fragen zu schreiben.

Main code	Subcode	Definition	Example	
			System	SCMC
	Enhance understanding	Tried to reduce the complexity of their production to enhance understanding.	Und, äh, vielleicht hatte ich [...] meinen Satz, ähm, zu kompliziert gemacht. (...) Weil [...] der Bot wollte, dass ich es bestätige (..) und dann probierte ich halt, es in einen Satz zu machen ((lacht))	Und wir verwendeten beide auch einfa-- einfache Sätze, damit wir es leichter kommunizieren können.
	Lost patience	Lost their patience to produce complex language because they were annoyed by the interlocutor.	Und eigentlich wechselte ich nie viel, also schrieb ich es einfach immer wie einfacher. Ich gab mir immer weniger um-- Mühe,	-
	Time pressure	Lack of complexity due to pressure to respond quickly.	-	ich versuchte schon, [...] mir Mühe zu geben, aber wenn man halt im Chat ist, hat man so, wie einen kleinen gewissen Zeitdruck und da überlegt man sich nicht (.) noch weitere Wörter oder irgendwie Synonyme für ein Wort (.) und darum, denke ich, machen wir es alle so einfach wie möglich.
	Etiquette	Sending additional messages before having received a response to the first message would not	-	bei Whatsapp ist es ja oft so, dass man etwas abschickt, dann fällt einem etwas ein, dann schickt man etwas

Main code	Subcode	Definition	Example	
			System	SCMC
		conform to chatting etiquette.		Neues und wieder etwas Neues. [...] aber wenn man halt etwas Geschäftliches schreibt, dann sollte man nicht jeden Gedanken einzeln so rausschicken, weil das könnte dann je nachdem auch nervig werden [...]

Note. Absence of an example in one interaction type (system, SCMC) means that the given subcode was not coded in the interview data of the respective group of participants.

Appendix M Codes pertaining to accuracy derived from learner interviews

Main code	Subcode	Definition	Example	
			System	SCMC
Focus	Spelling	The learner's statement focused on spelling accuracy.	von den Nomen [...] ob man sie gross schreibt am Anfang oder klein. (..) Das wusste ich nicht ((lacht)).	Und dann [...] begann ich mega, plötzlich an jedem Wort zu zweifeln. Ich dachte so, ist das jetzt wirklich so, kommt da noch ein "e" [...] weil ich wollte auch irgendwie keinen Fehler-- weil ich es der anderen Person halt auch wollte wie leicht machen.
	Lexis	The learner's statement focused on lexical accuracy.	[...] "mood" ist ja mehr so für "Stimmung" [...] "mood" ist eigentlich mehr so bezogen auf die Person (.) dann wusste ich nicht so recht, ja, wie soll ich das jetzt genau formulieren, aber ich wollte eigentlich wissen, wie, äh, die Stimmung ist in dem Raum.	da habe ich es nach-- [...] seine Telefonnummer und E-Mail wollte ich eigentlich. Und dann wusste ich nicht, wie man dem sagt, dann habe ich halt nur E-Mail-- dann merkte ich, dass es etwas komisch klingt mit "can you note me your e-mail" ((lacht)).
	Syntax	The learner's statement focused on syntactic accuracy.	-	Ich achtete, glaube ich, eh mehr auf die Satzstellung als dann auf die Rechtschreibung.
	General	The learner's statement focused on an unspecific aspect of accuracy such as the formulation of a complete room requirement.	[...] schrieb etwas gründlicher (...) und schaute auch auf-- ob ich Fehler habe oder nicht.	[...] ob ich es richtig formuliert habe und ob es auch wirklich die Antwort gibt, die ich möchte. Dann löschte ich es wieder und versuchte, in einem anderen Satz irgendwie besser

Main code	Subcode	Definition	Example	
			System	SCMC
				herauszuholen, was ich dann verantworten möchte.
Reasons for enhancing accuracy	Enhance understanding	Learner tried to increase the accuracy of their production to enhance understanding by the interlocutor.	[...] weil er "Beamer" nicht verstand. Und dann ging ich schauen, ob er schon einmal geschrieben hatte, [...] Projektor oder... wie das halt auf Deu... äh, auf Englisch heisst.	man muss überlegen, wie man das am besten schreibt, wie [...] wenn man mit jemandem telefonieren würde, [...] weil wenn man zum Beispiel telefoniert, hat die andere Person ja auch keine Zeit oder Lust, Rätsel zu raten, was man meint. Und dann muss man halt schauen, dass man so präzise und genau wie möglich das sagt, was man gerne hätte.
	Not specified	The learner does not specify why they tried to enhance the accuracy in their production.	Und wenn ich sonst schreiben würde, würde ich Dinge abkürzen, Wörter.	und dann habe ich noch einmal das Ganze durchgelesen und noch ein paar Fehler korrigiert, (.) die mir jetzt gerade aufgefallen wären.
Reasons for reducing accuracy	Lack of knowledge	Lacked the knowledge (i.e., adequate word, accurate spelling) to produce accurate language.	Dort war ja, dass man etwas auf dem Bildschirm oder so zeigen kann. Und ich wusste nicht, wie ich das schreibe.	wie man es halt auf Deutsch macht. Aber dafür habe ich etwas [...] Business Englisch Erfahrung, wie zum Beispiel, (..) ob er mir da eine Bestätigung schicken könne. Aber eben, da wusste ich nicht, wie man dem sagt.
	Speed	Work on accuracy would have impeded chatting speed.	ich schaute schon darauf, dass ich es richtig schreibe, aber jetzt--	irgendwie hatte ich das Gefühl auch, dass ich (..) manchmal

Main code	Subcode	Definition	Example	
			System	SCMC
			ich drücke jeweils schneller ab, als ich-- noch einmal-- ja, als ich noch einmal drüber nachdenke.	Dinge mega schnell abschickte. Also, das passiert mir auch sonst immer viel, wenn ich irgendwie-- wir haben bei uns im Geschäft Skype-- dass ich immer gleich sofort abschicke. Und dann im Nachhinein denke ich mir so, ah nein, und ist das jetzt wirklich so
	Mistyping	Lack of accuracy because of mistyping.	Ah, ich vertippte mich auch etwa hundert Mal, ja.	einfach nur weil ich oben eben das "g" vergessen hatte.
	Low-stakes exercise	The interaction was only a low-stakes exercise and not a graded test or a real-life conversation.	Das Datum, weil ich den Punkt machte und dann sendete ich es schon ab, ohne noch einmal-- ja, ich las es auch nicht mehrmals durch, weil ich dachte, ja, das ist ja nicht irgendwie ein Test	wenn ich ein Mail schreibe, dann lese ich es noch einmal durch-- das ich irgendwie im Unterricht abgeben muss oder so. Aber dort war halt so, ja, ich schreibe mal, dann schicke ich es ab. Dann schaue ich alles noch einmal durch, dann denke ich mir so, ist das wirklich richtig, habe ich es wirklich gut gemacht?
	No impact on understanding	The learner assumed that it would not affect the interlocutor's understanding.	Ich weiss nicht wirklich, ob man das auch auf Englisch machen kann, aber, ja (..) Dachte mir, die haben das ja sicher verstanden.	-
	Loss of patience	Learner lost their patience to produce accurate language	Wenn er nur alles in einem Sätze-- Satz geschrieben hätte und	-

Main code	Subcode	Definition	Example	
			System	SCMC
		because they were annoyed by the interlocutor.		wirklich eins nach dem anderen, dann hätte ich viel mehr dann weiter auch wirklich Mühe mir gegeben, um alles richtig zu schreiben. Aber es löschte mir so wie ein bisschen ab, ja ok, du bist ein Roboter, gut, ich schreibe dir einfach so zurück.
	Trial and error	Assumed that the interlocutor would let them know if they could not understand an utterance.	-	dann schrieb ich es einfach einmal und schaute, ob Antwort kommt mit "What do you mean?" oder ob sie es verstand.
	Lack of concentration	Learner lacked concentration, e.g., due to distractions such as noise in the classroom.	-	Und ja, wegen dem Konzentrieren, (..) ich glaube, alles war noch etwas laut im Raum, oder einfach noch ein paar am Reden, da konzentriert man sich halt auch nicht so-- nicht so gut.
	Cognitive overload	Learner was cognitively overloaded by producing the required content and writing accurately at the same time.	-	Ich machte, glaube ich, mega viele Fehler. (..) Ich weiss auch nicht, wieso, aber ich war auch mega-- also ich wusste nicht genau, wie soll ich jetzt anfangen, was soll ich jetzt machen. Waren so mega viele Fragen

Main code	Subcode	Definition	Example	
			System	SCMC
	Anxiety	Pressure to prevent mistakes and keep up with interlocutor resulted in decreased accuracy.	-	Ich hatte so das Gefühl, ja nichts falsch machen zu dürfen, und dann dadurch wirklich recht viel Zeit brauchte und auch Fehler machte und so.
	Chat modality	'Relaxed' character of chat modality led to decreased accuracy.	-	dass ich mir weniger Mühe gab, als jetzt bei einem Papierformat. Es macht es einem einfach recht gemütlich, wenn man am Chatten ist, und privat machen wir ja auch mehrere Fehler, (..) als jetzt in der Schule. [...]
	Not specified	The learner does not specify why their production lacked accuracy.	-	Ich achtete nicht so auf die Rechtschreibung.

Note. Absence of an example in one interaction type (system, SCMC) means that the given subcode was not coded in the interview data of the respective group of participants.

Appendix N Codes pertaining to communicative adequacy derived from learner interviews

Main code	Subcode	Definition	Example	
			System	SCMC
Selection strategy	Requested interlocutor's assistance	The learner requested the interlocutor's assistance, e.g., to present the available rooms with their characteristics or to suggest a suitable room based on the presented requirements.	[...] dem Computerprogramm jetzt zu erzählen, was ich haben muss, und sozusagen nach seiner Meinung zu fragen, was-- (.) also, (.) ob er mir helfen kann, einen Raum zu finden, wo (.) das alles zutrifft.	Ich versuchte einfach, ähm, zu fragen welches Zimmer es sein wird.
	One by one	Asked about the complete list of requirements one by one.	Also, dass ich mich einfach wie beim Blatt von oben nach unten ab--arbeitete [...] von den Kriterien, was es braucht, der Raum, und dann arbeitete ich einfach immer von oben immer nach unten alles ab.	Und dann hatten wir ja noch Punkte, [...] was das Zimmer haben sollte. Und dann versuchte ich so ein bisschen von dem aus, ähm, ja zu richten.
	Compare	The learner compared the information on the rooms against the requirements before selecting.	Also, da verglich ich ja jetzt die drei Räume (..) und (.) fand Alpha am passendsten [...] für mich der passendste Raum zu den Angaben. [...]	An dieser Stelle ging ich vergleichen mit dem ersten Raum [...] Konnte ich noch vergleichen gehen, während sie gerade schrieb. Dann schaute ich, welcher Raum besser ist im Auftrag.

Main code	Subcode	Definition	Example	
			System	SCMC
	Incomplete information	The learner selected the room before having collected all the required information.	und dann schrieb ich einfach, ja wenn er mich halt nicht versteht, soll er einfach Raum (.) Beta buchen.	-
	Personal prioritization	The learner selected a room based on personal prioritization of the requirements.	[...] es ging ja dann darum, wirklich, äh, die Informationen zu fragen und dann auszuwählen, welchen. Obwohl ich in mir selbst einfach wusste, ich möchte einfach den teuersten, weiss auch nicht, wieso.	-
	Excluded rooms	Reduced the selection by excluding rooms that appeared unsuitable.	-	wenn es hiess, okay, es hat keinen Projektor, dann wusste ich, okay, diesen Raum kann ich eigentlich gleich streichen. (.) Und-- oder wäre er irgendwie zu teuer gewesen oder so, hatte ich auch gl-- ok-- wusste ich auch gleich, okay, das ist es auch nicht.
	Followed interlocutor's suggestion	The learner followed the interlocutor's suggestion.	-	mir wurde der Raum vorgeschlagen und dann dachte ich mir, gut, weil sie wird ja wissen, was sie für Räume haben. Ähm, und da ich das ja nicht

Main code	Subcode	Definition	Example	
			System	SCMC
	Unsure	The learner was unsure about what strategy to follow.	-	entscheiden konnte, schrieb ich dann eben einfach "okay, gut" und fragte nach dem Preis. dass ich nicht so recht wusste-- ja, soll ich jetzt [...] also, nach diesen drei Zimmern zu fragen oder einfach das zu nehmen, das ich einfach gerade als Antwort bekommen habe.
Perception of task	Fun	The task was fun to do.	der Reiz ist halt mega schnell da, halt wirklich es auszuprobieren mit dieser Person zu schreiben, weil- - es ist halt schon etwas Spezielles. Es ist wirklich etwas, das macht man nicht jeden Tag und der Reiz ist wirklich da, halt so zu probieren, mit ihr zu chatten, also einfach zu chatten [...]	Aber ich fand es noch cool-- fand es eigentlich noch eine coole Abwechslung und so.
	Easy	The task was easy to solve.	und die Aufgabenstellung war auch klar. Und deshalb konnte man das eigentlich sehr gut lösen, finde ich.	Also, grundsätzlich war es nicht so ein schwieriger Auftrag.

Main code	Subcode	Definition	Example	
			System	SCMC
	Useful	The task was useful for learning, e.g., to try out communicating in practice or to focus on writing.	ich fand es wie cool, dass man mit jemandem, auch wenn die Person jetzt nicht echt war, kommunizieren konnte und das einmal wie ausprobieren.	Und dadurch, dass man quasi aus dem Englisch, das man hat, selber was zusammenstellen muss, oder musste, ähm, musste man halt auch einmal tiefer graben, als jetzt nur bei Tests, wo man genau das wissen muss, ähm, und dann musste man halt wirklich einmal alles, was man gegebenenfalls eben hat, herausholen, um das dann auch zu schreiben. (..) Und selbst wenn einem eben manche Wörter nicht einfelen, musste man dann halt nach einer Lösung suchen und ein anderes Wort dafür finden.
	Usefulness limited	The learner doubted the usefulness of the task for learning, e.g., because of lack of feedback or lack of practical relevance.	ich weiss jetzt nicht, was der jetzt-- also was das Programm jetzt von mir dachte und, verglichen mit den anderen, wo ich jetzt stehe. Vielleicht das ein bisschen, wo man steht. (.) Das wäre gut, ja.	[...] es war irgendwie so ein bisschen, ja, ich schreibe, aber ich weiss nicht wirklich, mache ich es jetzt richtig [...]
	Difficult	The task was difficult to solve, e.g., due to	ich wusste nicht mehr, eigentlich, genau, welcher	Ich hatte ein mega Durcheinander mit diesen

Main code	Subcode	Definition	Example	
			System	SCMC
		communication issues with the interlocutor or because it was hard to keep track of the interaction, specifically what information on which rooms have already been discussed.	"room"... weil ich eben eins nach dem anderen so ein bisschen fragte, wusste ich nicht mehr, welchen "room" ich schon gefragt hatte und welchen nicht. Und dann wegen der Zeit fragte ich einfach-- einfach wieder.	drei Zimmern. (.) Und da wollte ich eben auch fragen, welches davon es jetzt sein wird, damit ich mir nachher so wie einen Überblick machen kann, okay, welches würde ich jetzt nehmen?
	Interlocutor unreliable	The learner was uncertain if the interlocutor's information was reliable.	-	Ja, da war ich auch ein bisschen verwirrt ((lacht)), weil halt-- dann-- also vorher hatte sie ja wie gesagt, dass der einzige Raum von dem ich das Gefühl hatte, er geht, geht nicht. Und dann, musste ich wie noch einmal nachfragen, weil sonst für mich ist es-- also-- ich hatte dann wie einen Raum, bei dem ich gedacht hatte, der sollte ja eigentlich gehen, aber er hat keinen Beamer. Und dann fragte ich halt noch einmal nach, ob es <u>wirklich</u> so ist. (...) Weil es mir halt dann nicht so-- nicht so aufging.

Main code	Subcode	Definition	Example	
			System	SCMC
	Stressful	The learner felt anxious, e.g., about not knowing who the partner was or not wanting to make any mistakes.	-	also für mich persönlich war es so, zu wissen, dass es jetzt irgendjemand von der Klasse ist, hat mich fast wie, (.) auf eine Art, fast wie gehindert [...] Ich hatte so das Gefühl, ja nichts falsch machen zu dürfen, und dann dadurch wirklich recht viel Zeit brauchte und auch Fehler machte und so. Ich glaube, es wäre viel einfacher gewesen für mich, mit einem-- zu wissen, dass es jetzt irgendein Roboter ist, sozusagen, mit einem Computer zu schreiben.
Reasons for conformity to formal style	Appear professional	The learner wanted to appear professional.	ich darf ja nicht-- so ein bisschen professionell rüberkommen sollte ich.	-
	Email as model	To model the production after an email text.	ich wollte es eigentlich so führen wie ein-- wie ein Mail	-
	Not specified	The learner does not specify why they used a formal style.	also ich probierte selbst auch formell zu schreiben und auch ein bisschen im Business Englisch	Da suchte ich nach einem besseren Verb, damit es ein bisschen Business-mässiger klingt ((lacht)).

Main code	Subcode	Definition	Example	
			System	SCMC
Reasons for deviation from formal style	Mixed appropriate	Considered a less formal style as appropriate.	ich holte jetzt nicht mein Business Englisch raus. Im Betrieb [...] also ja, wenn ich ein-- einen Raum mieten würde, würde ich jetzt auch nicht mega, ultra Business Englisch sprechen. Darum dachte ich, eher so ein bisschen, ja, so Alltagsenglisch.	Also, zuerst dachte ich so, ja, soll ich es jetzt formell machen oder nicht? Weil grundsätzlich geht es ja schon an einen Bet-- also-- an einen Betrieb, sozusagen, aber dann dachte ich [...] das ist ein bisschen zu formell
	Lack of competence	Lacked the competence to produce formal style	Äh, ich hatte nicht wirklich eine professionelle-- ähm-- keinen professionellen Anfangssatz eigentlich (.) und dann dachte ich mir "Greetings from Switzerland"	ich wusste dann auch nicht, ob ich es durch das ganze Gespräch wie durchziehen kann. Weil man Business Englisch nicht so weit reicht.
	Unsure about appropriate style	Was unsure what style would be appropriate.	Da wusste ich nicht, wie ich anfangen sollte, das Gespräch. (..) Also, erst so simpel, ja, "Hallo" oder so wie wenn man ein Mail schreiben würde.	zuerst versuchte ich so ein bisschen-- halt so ein bisschen höflich alles zu machen, aber dann dachte ich, (.) nein, doch nicht und ich hatte oft so das Gefühl, soll ich das jetzt-- [...] ich dachte so, muss ich es formell machen, [...] ja, ich war mega, eben, unsicher und wusste nicht wirklich, was machen. Weil ich wollte auch nichts irgendwie falsch

Main code	Subcode	Definition	Example	
			System	SCMC
				machen, aber ich musste ja doch irgendetwas machen.
	Not specified	The learner does not specify why their style deviated from formal style.	und dann, als ich immer mehr im Text war, bin-- schrieb ich immer weniger professionell.	dass man eben zum Beispiel einmal vergisst, den Anfang gross zu schreiben oder eher so privat-mässig schreibt, also nicht wirklich so geschäftlich, sondern-- ich weiss nicht, wie ich das erklären soll-- einfach so, wie man auch eben mit Freunden zum Beispiel schreiben würde.
	Interlocutor's behavior	Chose a less formal style because of the interlocutor's irritating behavior.	da löschte es mir wohl etwas ab, eben weil ich-- weil ich das Problem hatte mit dem Roboter [...] Deswegen schrieb ich dann "yeah". "Yeah, what type", das war dann schon gar-- gar nicht mehr professionell.	-
	Classmate as interlocutor	Chose a less formal style because the interlocutor was a classmate rather than an actual hotel employee.	-	also ich hatte ja jetzt eben eine Person aus unserer Klasse und ja, (.) da wusste man ja, eben, dass es die-- irgendeine Person davon war. Und dann war es halt noch einmal etwas

Main code	Subcode	Definition	Example	
			System	SCMC
	Habit	The learner chose a less formal style because it corresponds to their usual chatting habit.	-	anderes, dass man eher wieder in diese Gewohnheit zurückfällt. dass es ja eigentlich mehr in Richtung normales-- auch ein bisschen wie Business-mässig eigentlich sein sollte, aber dass man automatisch wieder in sein normales Chat-Verhalten reinfällt, wie man es halt auch auf Whatsapp oder so macht.
	Enhance understanding	The learner chose a less formal style on to enhance understanding.	-	[...] versuchte immer, so ein bisschen einfach alles zu formulieren, damit es für die andere Person möglichst einfach ist und für mich auch möglich einfach, damit die andere Person dann auch nicht mega formell oder so schreibt.
Behaviors enhancing conformity to generic structure	Stay on topic	The learner made sure to stay on topic, e.g., by not switching topics abruptly or ensuring the relevancy of their messages.	[...] eben nicht irgendwie vom einen Thema immer zu verschiedenen Themen zu wechseln, also immer systematisch vorzugehen.	ich wollte eigentlich noch etwas Weiteres schreiben, (.) aber ich habe es dann gemerkt, dass es dann irgendwie nachher (.) keinen Sinn ergeben würde, oder ja, (.) dass es

Main code	Subcode	Definition	Example	
			System	SCMC
				eigentlich nebensächlich ist.
	Manage interlocutor deviations	Resisted interlocutor deviations from the generic structure, e.g., by ignoring non-relevant interlocutor messages or by negotiating meaning.	Ich verstand dann auch nicht ganz, wieso er mir, danach, [...] als ich ihm sagte, welchen Raum ich will, dass er mir noch einmal sagte, dass der auch, ähm, Bildschirme hat und Lautsprecher, weil das hatte er mir ja eigentlich schon vorher gesagt. Aber das habe ich dann auch-- bin ich nicht mehr wie drauf eingegangen, weil es ja eigentlich gut ist, dass er es mir noch einmal sagte.	Ähm und dann fragte sie ja-- die Person (unv.) das war, direkt, welchen Raum ich gerne hätte. Und dann fing ich halt erst an zu schauen wegen der Grösse,
	Wait for interlocutor messages	The learner waited for their interlocutor to respond in order to adapt their message to the interlocutor's, e.g., to ensure its relevancy.	-	erst einmal, welche Räume er mir aufzählt, die genau Platz hätten. Und dann hätte ich-- wollte ich, glaube ich, schauen, welche Frage am besten (..) äh, dann dazu passt. Nicht, dass ich dann noch unnötige Fragen stelle.
Behaviors diminishing conformity to generic structure	No strategy	The learner did not follow a specific strategy when writing, e.g., because they	[...] Und dann begann ich einfach zu schreiben und man merkte-- also-- aber	[...] Und dann wusste ich nicht, ja, was soll ich jetzt als erstes fragen und so.

Main code	Subcode	Definition	Example	
			System	SCMC
		were unsure about what structure would be appropriate.	ich weiss einfach, dass es-- ich schrieb einfach, wie ich es schreiben würde	Und dann, ja, fing ich einfach einmal an, so ein bisschen zu schreiben
	All at once	The learner provided all information at once.	[...] dann entschied ich mich eben, so wie ein Mail-- also erst einmal, (..) ähm, viel Text (..) und nicht-- äh-- und dann gleich alle Informationen so	wenn man halt etwas Geschäftliches schreibt, dann sollte man nicht jeden Gedanken einzeln so rausschicken [...] sondern sich einmal überlegen, was man sagen möchte und das dann auch so zusammenfasst, dass es <u>eine</u> Nachricht, wo alles-- also jetzt klar, nicht alle Sachen drin sind, aber-- nicht einzelne Gedanken.
	Guided by interlocutor	The learner's behavior was guided by the interlocutor.	und dann gaben sie mir schon wie den Anstoss so was und wie-- gab eben eine Idee, wie ich eben starten soll.	Es dünkte mich zwar am Anfang etwas schwierig so mit dem Anfangen, weil ich nicht wusste, ob ich jetzt anfangen muss oder nicht. Weil, es wäre ja eigentlich meine Fragen gewesen, weil ich wollte wissen, wie die Räume aussehen. Da war ich fast froh, dass sie zuerst anfing. Da ging es nachher einfacher.

Main code	Subcode	Definition	Example	
			System	SCMC
	Extended conversation	Extended the conversation beyond the required topics of the task, e.g., to test out the limits of the dialog system.	Habe ich eben mal gefragt, ja, "und wie geht's dir?" und so, zum Schauen, und ja--	-
	Need for restart	Needed to start the conversation again from the beginning due to the interlocutor's non-understanding.	ich war dann etwas verwirrt, als ich sagte, ich habe noch eine Frage und er sagte einfach, "Danke, dass Sie uns ausgewählt haben", und dann habe ich einfach noch einmal wie von Anfang angefangen, was überhaupt der Grund ist, warum ich ihm schreibe. Also, ich habe dann noch einmal geschrieben, dass ich jetzt gerne noch den Raum reservieren und buchen würde	-

Note. As a note on the aspect *style*, although learners generally did not explicitly mention “style” in their statements, the relation could be deduced from the context. For instance, in the example for the code *appear professional* was further explained by the learner in the subsequent response, where it became clear that the learners’ first statement related to the formality of the greeting. Absence of an example in one interaction type (system, SCMC) means that the given subcode was not coded in the interview data of the respective group of participants.