

## Editorial

**Annelies Häcki Buhofer (Basel)**

Zum Konzept des Korpus gehört – nicht unabdingbar, aber usuell und je länger je mehr – eine Grösse, die maschinell durchsucht werden muss und "von Hand" bzw. korrekter mit dem lesenden Kopf – oft auch in Auszügen oder in durch einfache Suchprozesse erstellten umfangreiche Listen – nicht mehr bewältigbar ist.

In dieser Situation sind Suchstrategien – ihre Entwicklung und Beurteilung – nicht nur für das Finden des interessierenden "Materials im Material" von Bedeutung, sondern auch für die Beurteilung der Brauchbarkeit des Korpus. Wenn sich ein gesuchtes Phänomen auf Anhieb im Korpus nicht finden lässt, ist zunächst offen, was daraus zu schliessen ist: Das kann ebenso gut bedeuten, dass die Suchstrategien untauglich sind, wie es heissen kann, dass das gesuchte Phänomen für die Texte des Korpus nicht von Bedeutung ist, dass das Fehlen eines Phänomens entsprechend charakteristisch ist für die Zeitabschnitte oder die Textsorten, die in einem Korpus repräsentiert sind. Beide Aspekte sind also theoretisch zu analysieren: Die Such- und Filterstrategien ebenso wie das Verhältnis von Quantität und Qualität eines Korpus. Nicht immer jedoch findet man (zu) wenige der gesuchten Phänomene – manchmal findet man auch (zu) viele und es ist dann die Frage, wie die Ergebnisse auf die wirklich interessierenden Phänomene hin gefiltert werden können. Die Filterprozeduren sortieren die unbrauchbaren Suchergebnisse in eine Negativliste aus. Wenn die Positivliste "nichts hergibt", so ist das im obigen Sinne aussagekräftig für das zu untersuchende Phänomen, wenn man zuverlässig und begründet ausschliessen kann, dass das Korpus einseitig und/oder schlecht strukturiert ist und dass die Such- und Filtermethoden suboptimal geeignet sind, ein Phänomen anzutreffen.

Welche Arten von Vorüberlegungen man sich machen muss bzw. was man sich einfallen lassen muss, um wirklich zu finden, was man sucht, zeigen an unterschiedlichen Fragestellungen und Korpora die Beiträge zu folgenden Fragen dar: Wie kann man ein relativ abstraktes Phänomen wie die indirekte Rede finden? (A. Hornung), wie kann man die Paraphrasentexte eines Wörterbuchs als Korpus nutzen, um historische Phraseologismen zu finden, die als Lemmata nicht aufgeführt sind? (M. Dräger), und wie können unbekannte und lexikographisch interessante Wörter aus grossen Textmengen extrahiert werden? (A. Geyken). Die Antworten lassen sich nicht leichthin geben, sondern stellen eigene Forschungsbeiträge dar.

Aber auch der Begriff des Korpus muss überdacht werden: Kann man von einem nicht-aufgenommenen Korpus mündlicher Äusserungen sprechen? (Ch. Rittaud-Hutinet). Die mündlichen Texte sind im Kontext der Korpuslinguistik noch ausgeprägter in der Minderheit und damit relativ zu schriftlichen Texten vernachlässigt, als das in der Fachgeschichte schon immer der Fall war, weil es vorderhand immer noch so unendlich aufwändig ist, sie zu transkribieren und damit erst die Voraussetzungen für die schriftliche Durchsuchbarkeit zu erstellen, die aber wegen der unterschiedlichen Aussprachen und allen weiteren artikulatorischen Besonderheiten der gesprochenen Sprache ungleich viel anspruchsvoller bleibt.

Die Fortschritte der Korpusarbeit zeigen sich auf verschiedenen Ebenen: Zunächst einmal mit Bezug auf das Korpus selber, dessen Quantität durch möglichst viel opportunistische Anteile wächst, dessen Qualität jedoch auf Strukturierung und definierte Auswahl angewiesen ist (H. Bickel, M. Gasser, A. Häcki Buhofer, L. Hofer, Chr. Schön). Die Fortschritte zeigen sich jedoch auch weiter mit Bezug auf die Differenzierung der Aufgaben, bei denen ein Korpus

und eine entsprechende Datenaufbereitung hilfreich sein kann, um Rohdaten für eine Reihe von weiteren linguistischen Aufgaben zu erzeugen (U. Quasthoff).

Korpora stellen auch nicht bloss eine klar definierte und abgegrenzte Sprachbasis zur Verfügung, sondern können auch als Schatzkästlein dienen. So bei bedrohten Sprachen, für die ein Korpus ein Denkmal mit kulturpolitischer Zielsetzung sein kann (R. Marti, B. Andreeva, W. Barry).

Die korpusbasierten Studien ermöglichen sprachstrukturell und vergleichend vertiefte Einsichten, wie einige der vorliegenden Studien zeigen: Die Frage nach den Unterschieden in englischen und deutschen Nominalphrasen und der Ausprägung dieser Merkmale in Fachtexten und deren Übersetzungen lässt sich nur mit mehreren angemessenen Korpora beantworten (S. Hansen-Schirra, S. Hansen, S. Wolfer, L. Konieczny). Ebenso kann die Frage der Häufigkeit von Unterbedeutungen und Kookkurrenzpartnern eines Verbprototyps wie einstellen in verschiedenen europäischen Sprachen nur anhand von mehreren Korpora angegangen werden (R. Duffner, A. Kamber, A. Näf).

Die schwierigen Fragen und harten Nüsse der Korpus- und Computerlinguistik werden auch nicht ausgespart und beispielsweise an den Phraseologismen und Formen der Sprachgeschichte bearbeitet (N. Filatkina). Überhaupt stellt die Sprachgeschichte einen Prüfstein der Korpuslinguistik – und der Korpora selber – dar (D. Elmiger).

Dass ein definiertes Korpus eine seriöse Grundlage einer empirischen Untersuchung darstellt, ist in der Sprachwissenschaft – je nach Ausrichtung – seit längerem oder seit langem bekannt. Wie digitale Korpora über die opportunistische Erhältlichkeit von Daten hinaus ausgewogen und strukturiert erstellt werden können, und wie – mit welchen Such- und Filtermethoden – sie dann für die verschiedensten Fragestellungen möglichst optimal genutzt werden können, stellt eine vielschichtige Thematik dar, die in *Linguistik online 39, 3/2009* unter dem Titel "Fortschritte in Sprach- und Textkorpusdesign und linguistischer Korpusanalyse II" von verschiedenen Standpunkten aus angegangen wird.

Universität Basel im Oktober 2009