

Editorial

Annelies Häcki Buhofer (Basel)

Die Korpuslinguistik erlebt einen eigentlichen Boom. Korpora als Grundlagen für linguistische Fragestellungen, deren Beantwortung grössere und vielfältigere Datenmengen erfordert, als man als Einzelperson sammeln und analysieren kann, sind im Trend. Eine Reihe von EinzelforscherInnen, Forscherteams und Institutionen stellen sich der Herausforderung, Wege zu finden zwischen einer angemessenen Strukturierung eines Korpus, der Berücksichtigung der vielfältigen Texte einer geschichtlichen Periode, einer brauchbaren Grösse und der Beschränkung personeller Ressourcen, die durch die Automatisierung von Arbeitsschritten gemildert, aber bei weitem nicht gänzlich aufgefangen werden kann.

Diese Hoch-Zeit der Korpora scheint vorderhand ungebrochen – obwohl die Pflege und Erhaltung der Korpora in den wenigsten Fällen gesichert ist. Da Forschungsprojekte im allgemeinen für eine sehr begrenzte Zeit und Aufgabe finanziell unterstützt werden können, müssten es die Institutionen sein – die Akademien, nationalen Forschungsförderungsinstitutionen oder grossen Sprachinstitute, die diese Aufgabe übernehmen oder platzieren. Dass eine kommerzielle Nutzung von Korpusgrundlagen die Weiterentwicklung oder auch nur die blosser Erhaltung und technische Pflege ermöglicht, bleibt wohl die Ausnahme. Dass die europäischen Urheberrechtsgesetze dem nicht-kommerziellen, wissenschaftlichen Interesse an umfassenden Textsammlungen so wenig entgegenkommen, wie sie das derzeit tun, bleibt vorderhand ein Mangel.

Während die schiere Grösse und Reichhaltigkeit der Ressourcen eines Korpus bisher und nach wie vor seine Bedeutsamkeit begründen (Institut für Deutsche Sprache, www.ids-mannheim.de), wird zunehmend auch die Strukturierung eines Korpus ein wichtiges Qualitätskriterium (Digitales Wörterbuch, www.dwds.de, Schweizer Textkorpus, www.schweizer-textkorpus.ch), ein zunehmend sorgfältiger Aufbau also, der nicht nur berücksichtigt, was leicht erhältlich ist oder digital vorliegt, sondern auch kleinere Texte und Textmengen aus entlegeneren oder sozial oder thematisch spezialisierteren Kontexten miteinbezieht. Weiter werden Korpora zunehmend an ihrer historischen Tiefe gemessen sowie daran, für wen sie in welcher Weise nützlich oder interessant sind. Nicht alle Korpora können vorderhand – oft aus urheberrechtlichen Gründen – extern abgefragt werden (Austrian Academy Corpus, www.aac.ac.at).

Dass das Interesse an Sprachdaten, gerade wenn auch einige interessante Auswertungsmöglichkeiten der Korpora angeboten werden, auch von Seiten nicht-fachwissenschaftlich interessierten Nutzerinnen und Nutzern gross sein muss, zeigen die Zahlen des Projekts Deutscher Wortschatz (wortschatz.uni-leipzig.de). Die grossen zukunftssträchtigen Themen sind automatisiertes Suchen und Erstellen von Rohdaten, die Kombination von quantitativen Möglichkeiten und bisher nur qualitativ beantwortbaren Fragestellungen und die bessere Nutzung der Korpora-Möglichkeiten für differenzierte Fragestellungen.

Die Korpuserstellungspraxis ebenso wie die Korpustheorie hat sich in verschiedenen linguistischen Teilgebieten mit nicht eben viel wissenschaftlichem Austausch entwickelt. Die methodologischen Entwicklungen von CHILDES für die Kindersprachforschung, die Korpustheorie, die sich begleitend entwickelt hat, ist von der textorientierten Datenbankentwicklung nicht genutzt worden und umgekehrt. In der breiten Nutzung von Erfahrungen liegt ein Potential, das es in Zukunft besser zu nutzen gilt – auch zur ausgeprägteren Differenzierung von Korpora, die durch ihre Funktionen als Referenzkorpora für zeitlich, regional und sozial be-

stimmte Texte dialogischer oder monologischer Natur von Kindern oder Erwachsenen etc. genauer bestimmt werden können.

Universität Basel im März 2009 – Annelies Häcki Buhofer