

Vorwort

Das *Kompendium Korpuslinguistik*¹ hat zum ersten Mal einen Überblick über Forschungsmethoden und aktuelle, vor allem deutsch-tschechische Projekte geboten. In den theoretischen Beiträgen zu diesem Band (vor allem Kratochvílová/Wolf und Eichinger) wurde ein spezielles Problem allen sprachwissenschaftlichen Arbeitens angesprochen: Jedes Projekt braucht als Grundlage Daten; dies dürfte unstrittig sein. Die entscheidende Frage aber ist die, wie man zu sprachlichen Daten kommt. Im Gegensatz etwa zur Geschichtswissenschaft hat die Sprachwissenschaft ihre Quellen für die zu interpretierenden Daten bislang nur ansatzweise reflektiert; eine ‘sprachwissenschaftliche Quellenkunde’ steckt allenfalls in den Kinderschuhen.

Dennoch hat sich mittlerweile eine ‘Korpuslinguistik’ etabliert, die bestrebt ist, einerseits auf ‘Korpora’ als einen Typus möglicher Datenquellen aufmerksam zu machen, ‘Korpora’ als Datenquellen zu definieren; andererseits arbeitet ein Teil der ‘Korpuslinguisten(-innen)’ daran, große Textkorpora mit allem, was dazu gehört, zusammen- und der wissenschaftlichen Gemeinschaft zur Verfügung zu stellen. So gesehen, wird unter einem ‘Korpus’ eine feste Sammlung von – häufig zahlreichen – Texten verstanden, die maschinenlesbar zur Verfügung steht und mit Annotationen und Meta-Daten versehen ist. Die Technik erlaubt es, die gewonnenen Daten ‘exhaustiv’, ‘frequenzorientiert’ und ‘kontextsensitiv’ zu analysieren und zu interpretieren. Ein Problem taucht bei jedem korpuslinguistischen Arbeiten auf: Es wird nur das analysiert und beschrieben, was sich im Korpus findet. Unter Umständen theoretisch hochinteressante Phänomene, die in einer Belegsammlung wie dem alten Zettelkasten als Lese Früchte gespeichert waren, können im Korpus nicht auftauchen.

Zudem hat Hans Wellmann im *Kompendium* gezeigt, dass ein Wörterbuchmacher ein Stichwort oder eine Bedeutungsvariante nicht in ein (Lerner-)Wörterbuch aufzunehmen hat, selbst wenn es in einem Korpus häufig belegt ist. Mit anderen Worten, die Belegfrequenz kann nicht das einzige Kriterium für eine bestimmte Interpretation oder Hypothese sein. Glücklicherweise ist der sprachwissenschaftliche Sachverstand (noch?) nicht ersetzt – und wird es unserer Meinung nach zukünftig auch nicht werden.

¹ Kratochvílová, Iva/Wolf, Norbert Richard (Hg.) (2010): *Kompendium Korpuslinguistik*. Eine Bestandsaufnahme aus deutsch-tschechischer Perspektive. (= Germanistische Bibliothek 38). Heidelberg .

Es hat sich gezeigt oder bestätigt, dass für jede Untersuchung ein eigenes, ein spezielles Korpus zugrunde gelegt werden muss. Selbst wenn man auf eines der großen öffentlichen Korpora zurückgreift, muss dies gerechtfertigt werden. Andererseits ist auch immer zu prüfen, welche Texte in welcher Form in ein Korpus aufgenommen werden.

Immer wieder muss geprüft werden, ob ‘Große’ und ‘Kleine Korpora’ als Korpora brauchbar sind und auch kombiniert werden können/müssen. Beim derzeitigen rasanten technischen Fortschritt stellt sich stets aufs Neue die Frage, mit welchen technischen Mitteln die Daten gespeichert und welche Meta-Daten und Annotationen angebracht werden sollen. Auch die Frage ‘korpusbasiert’ vs. ‘korpusgeleitet’ muss immer wieder zur Sprache kommen.

Bereits im Kompendium aber finden wir Ansätze zu einem erweiterten Korpus-Verständnis: Sowohl Sprach-/Dialektatlanten als auch Datenbanken sind unter bestimmten Bedingungen als Datenquellen für korpuslinguistisches Arbeiten geeignet. Es ist dies ein Ansatz, der jetzt von Wolf Peter Klein in verstärktem Maße im Sinn einer ‘Datenbanklinguistik’ aufgegriffen wird.

Mit anderen Worten: Auch wenn das *Kompendium* bereits viele Fragen beantwortet, es wirkt in erster Linie doch anregend und nicht erschöpfend. Deshalb haben sich zum zweiten Mal zahlreiche Sprachwissenschaftler(-innen) aus der Tschechischen Republik und aus Deutschland vom 25. bis 27. Oktober 2011 zur ‘II. Internationalen Konferenz „Korpuslinguistik Deutsch-Tschechisch kontrastiv“’ im Bildungshaus Sambachshof und an der Julius-Maximilians-Universität Würzburg zusammengefunden, um gemeinsam Antworten auf die grundsätzliche Frage nach den ‘Grundlagen einer sprachwissenschaftlichen Quellenkunde’ zu erarbeiten. Die Antworten sind so vielfältig wie die zugrundeliegenden Projekte und Forschungsaufgaben.

Ermöglicht wurden diese Zusammenkunft und die gemeinsame Arbeit durch die Gemeinnützige Hermann-Niermann-Stiftung, Düsseldorf, ohne deren Munifizienz so vieles nicht möglich geworden wäre. Die spezielle Atmosphäre im Bildungszentrum Sambachshof (bei Bad Königshofen) schafft überdies die Grundlage für gedeihliches Arbeiten und kollegiale Gespräche.

Dem Institut für Deutsche Sprache in Mannheim, das in der Korpuslinguistik eine wichtige Rolle spielt, danken wir für die Aufnahme dieses Bandes in die renommierten *Schriften zur deutschen Sprache* und für vielerlei Hilfe.

An beiden deutsch-tschechischen korpuslinguistischen Konferenzen hat Hans Wellmann teilgenommen und sie nicht nur durch seine fachliche Kompetenz,

sondern viel mehr noch durch seine einzigartige Fähigkeit, auf jeden Menschen kollegial und freundschaftlich zuzugehen, und durch seinen feinen Humor ganz wesentlich geprägt. Am 1. März 2012 ist Hans Wellmann verstorben. Er hinterlässt eine Lücke, die wohl kaum zu füllen sein wird. Aus vielen Gründen werden wir ihn nicht vergessen; für den vorliegenden Band hat uns Jana Valdová, Hans Wellmanns Frau, das Manuskript seines Vortrags zur Verfügung gestellt. Dieser Beitrag ist Hans Wellmanns letzter Beitrag für unser Fach; er hat, nicht nur für uns, den Charakter eines Vermächtnisses. Es ist uns eine Ehre und eine Verpflichtung, dass wir diesen Beitrag veröffentlichen dürfen. Der vorliegende Band sei dem Andenken an Hans Wellmann gewidmet, den keiner von uns vergessen wird.

Würzburg und Opava
im April 2013

Iva Kratochvílová
Norbert Richard Wolf