

Bruno Strecker (Mannheim, Deutschland)

Korpusgrammatik zwischen reiner Statistik und intelligenter Grammatikografie

PLENARVORTRAG/KEYNOTE SPEECH

Abstract

Klassische Grammatikschreibung geht – teils mehr, teils weniger bewusst – davon aus, dass dem, was an sprachlichen Äußerungen anzutreffen ist, ein mehr oder weniger homogenes System von Kategorien und Regeln zugrunde liegt, das in einer Art Simulation der zu erfassenden Sprache expliziert werden kann, derart, dass idealerweise die sprachlichen Hervorbringungen von Personen, die sich an die Simulation halten, von jenen anerkannt kompetenter Sprecher der Sprache nicht zu unterscheiden wären. Als empirisch kann diese Art der Grammatikografie deshalb allenfalls insoweit gelten, wie sie ihre Mutmaßungen über die Sprache einer kritischen Bewertung durch unstrittig kompetente Sprachteilhaber aussetzt.

Solange Grammatiker sich im Wesentlichen nur auf ihre Sprachkompetenz stützen konnten, gab es zu diesem Vorgehen keine echte Alternative. Durch die Möglichkeit, riesige Textkorpora maschinell zu durchforsten, haben sich für Grammatiktheorie wie für praktische Grammatikforschung jedoch neue Perspektiven eröffnet:

- Aus sprachtheoretischer Sicht kann die Frage nach dem Gegenstand der Grammatik neu beantwortet werden, denn an die Stelle so problematischer Hypostasierungen wie „Sprache“, „Sprachsystem“ und „Einzelsprache“ kann jetzt eine überwältigende Fülle maschinell auswertbarer Einzelerfahrungen eben jener Art treten, auf denen letztlich jede Sprachkompetenz aufbauen muss. Fragen wie jene, was das Deutsche eigentlich sei, beantworten sich auch jetzt nicht von allein, doch sie verlieren viel von ihrer scheinbaren Tiefgründigkeit und stellen sich neu als überschaubares praktisches Problem, nämlich eine geeignete Datenbasis zusammenzustellen.
- Die Forschung kann sich endlich an die Wittgensteinschen Maxime halten, nicht immer gleich auszudenken, was der Fall sein könnte, sondern erst einmal nachzuschauen.

Die Grammatikforschung kann sich radikal auf die neuen Möglichkeiten einlassen und ganz mit der Tradition brechen, indem sie alle Mutmaßungen über Bord wirft, die auf Individualkompetenzen gestützte Grammatiken über die Struktur ihrer hypostasierten Objekte angestellt haben, und dann Wege suchen, Grammatiken im Zuge geeigneter statistischer Auswertung der verfügbaren Daten gleichsam von selbst entstehen (emergieren) zu lassen.

Weniger radikal, mehr der Tradition verpflichtet geht das Projekt *Grammatische Variation im standardnahen Deutsch* vor, das hier vorgestellt wird. Es greift durchaus auf, was Generationen von Grammatikern zusammengetragen und festgehalten haben. Es sieht in der Korpusrecherche vor allem ein Mittel, Aussagen zur Grammatik des Deutschen zu evaluieren. Unabdingbare Voraussetzung für die Umsetzung dieses Vorhabens ist natürlich eine Datenbasis, die als ausgewogen und repräsentativ dafür gelten kann, was Menschen, die sich als deutschsprachig verstehen, an sprachlichen Äußerungen begegnet. Dass schon dies enorme Schwierigkeiten bereitet, versteht sich von selbst und wird sicher ein Hauptgegenstand der Diskussion sein, denn genau genommen nimmt die Entscheidung für die Textbasis all die Fragen über Grammatikalität und Akzeptabilität vorweg, die sich Grammatikern immer schon, doch eben punktuell bei ihrer Arbeit gestellt haben. Doch, wie immer man sich entscheidet, in jedem Fall wird offen zu Tage liegen, was Gegenstand der Untersuchungen sein soll. Das Korpus wird dann Ausgangspunkt sein

- für eine Evaluation der Geltung von Aussagen über grammatische Regularitäten, die in wissenschaftlichen Grammatiken aufgestellt wurden (also etwa von Aussagen zum Konjunktivgebrauch, zur Rektion, zur Wortstellung, Wortbildung und zur Existenz paradigmatisch zu erwartenden Wortformen),
- für die Beurteilung von Varianten, Zweifelsfällen und Normverstößen, die nicht als singuläre Fehlleistungen abgetan werden können (also etwa *wir Deutsche* oder *wir Deutschen*, *dem Opa seine Brille*),
- und, wo und soweit dies möglich scheint, für Erklärungen von Erscheinungen der genannten Art anhand frequenter, als alternativ erkannter Muster (also etwa *diesen Jahres* analog zu *vergangenen, kommenden, folgenden Jahres*).