

Einrichtung eines Forums

Die *Zeitschrift für Sprachwissenschaft* möchte ab dieser Nummer regelmäßig ein Forum zur Diskussion linguistischer Fragestellungen anbieten. Das Ziel dieses Forums ist es, durch ein möglichst breites Spektrum von unterschiedlichen Perspektiven bisher zu wenig beachtete oder kontrovers diskutierte linguistische Konzepte, Methoden und Phänomene genauer zu beleuchten und damit linguistische Zusammenhänge auch über Theoriegrenzen hinweg vorzustellen und Anregungen zu neuen Lösungswegen zu geben.

Zum Auftakt werden in dieser Nummer die Diskussionsbeiträge zum Thema „Linguistische Daten: Erhebung – Beurteilung – theoretische Implikationen“ abgedruckt, das vom Redaktionsteam formuliert wurde. In Zukunft sollen Vorschläge von unseren Leserinnen und Lesern zum Zuge kommen. Wer ein Thema vorschlagen möchte, sende einen zwei- bis dreiseitigen Text mit einer kurzen Beschreibung des Themas und den Problemen seiner Erfassung, ggf. auch eine durchaus kontrovers zu diskutierende Hypothese, an das Redaktionsteam (Termin und Adresse s. unten), welchem die Entscheidung über das zu wählende Thema obliegt. Nach der Ankündigung des Themas nimmt die Redaktion kritische Beiträge im Umfang von drei bis fünf Seiten entgegen, die in der darauf folgenden Nummer publiziert werden. Die Ankündigung des nächsten Forum-Themas wird in Heft 28.2 (2009, Nr. 2) erfolgen, so dass die konkreten Beiträge im Heft 29.1 (2010, Nr. 2) abgedruckt werden können. Die Annahme von Beiträgen wird vom Redaktionsteam zusammen mit den Verfassern des Forum-Themas entschieden.

Forum-Themen für das nächste Forum sind bis zum **31. Juli 2009** zu richten an:

wbisang@mail.uni-mainz.de

Walter Bisang
Johannes Gutenberg-Universität Mainz
Department of English and Linguistics (FB 05)
Jakob-Welder-Weg 18
D-55099 Mainz

Das Redaktionsteam der *Zeitschrift für Sprachwissenschaft*

Linguistische Daten:

Erhebung – Beurteilung – theoretische Implikationen

Sprache ist durch ihre Variabilität gekennzeichnet – oder wie kein geringerer als Sapir (1921: 147) gesagt hat “Everyone knows that language is variable”. Linguisten aus den unterschiedlichsten Richtungen suchen nach Verallgemeinerungen und strukturellen Zusammenhängen in dieser dem Wandel unterworfenen Vielfalt. Zu diesem Zweck greifen sie mit unterschiedlichen Methoden auf Datenausschnitte zurück, die sie zur Stützung ihrer Abstraktionen verwenden. Dass die Art der Daten Auswirkungen auf die Theoriebildung haben kann, wird dabei oft übersehen bzw. relativ selten explizit thematisiert (s. aber SFB 441: „Linguistische Datenstrukturen“, Tübingen).

Die mittlerweile schon als klassisch zu bezeichnende Unterscheidung zwischen auf Introspektion beruhenden, elizitierten Daten und korpusbasierten Daten mag dies belegen. Aus formallinguistisch-syntaktischer Sicht ist das kognitiv-kategorielle sprachliche Wissen und damit die Universalgrammatik (UG) primär über elizitierte Sprachbeispiele zugänglich, die auf der Intuition des linguistischen Forschers und auf der Intuition anderer Sprecher basieren. Sprachkorpora dagegen geben, auch wenn sie noch so umfangreich sind, die zur Ermittlung der UG notwendigen extrem seltenen Beispiele nicht oder nur teilweise her, enthalten keine negative Evidenz und, wie etwa Newmeyer (2003) betont, liefern sie lediglich Fakten über den (externen) Sprachgebrauch einer Gemeinschaft und lassen somit keine Rückschlüsse auf das grammatische Wissen eines Individuums zu.

Die Wertungen für Grammatikalitätsurteile bei auf Intuition beruhenden Beispielen reichen von einer strikten Dichotomie [\pm grammatisch] über [\pm grammatisch] plus [optional] bis hin zu einer sieben Punkte umfassenden Akzeptabilitätsskala bei Belletti & Rizzi (1988). Während sich einfache Wertungssysteme oft als zu grob für die Erfassung von Spre-

cherurteilen erweisen, werden komplexere Systeme wie das eben genannte als inkonsistent bezeichnet. So kritisiert Schütze (1996: 43) an diesem System u. a., dass eine allgemeine Theorie über Prinzipien, die stärkere bzw. schwächere Grammatikalitätsverletzungen auslösen, fehlerhaft und dass die Beurteilung von konkreten Verletzungen *ad hoc* sei. In einem vor kurzem erschienenen *Target*-Paper zur Datenerhebung in der generativen Linguistik spricht Featherston (2007) explizit von unterschiedlichen Graden der Wohlgeformtheit und fordert die Überprüfung sprachlicher Daten durch das Urteil einer größeren Zahl von Informanten (20 – 30) und mit verschiedenen lexikalischen Varianten. Mit diesem Ansatz, der sicherlich eine erhebliche Verbesserung der Datenqualität mit sich bringt, bleibt allerdings das Problem, wie ein Kontinuum von Urteilen mit der Grammatikalität korreliert, besonders dann, wenn davon auszugehen ist, dass bei dieser Methode eine Vielzahl von außerlinguistischen Faktoren mit in die Urteile einfließen.

Aus korpuslinguistischer Perspektive ist die Kritik an der Validität von auf Intuition basierten Abstraktionen nicht weniger stark. Zur Illustration sei hier Manning (2003: 296) zitiert:

[G]enerative grammar has produced many explanatory hypotheses of considerable depth, but is increasingly failing because its hypotheses are disconnected from verifiable linguistic data.

Nach Mannings (2003) Ansicht postulieren generative Theorien einerseits zu viel, weil sie statistisch seltene Fälle übersehen, die sich mit Korpora entdecken lassen, und andererseits erklären sie zu wenig, weil sie nichts über weiche Beschränkungen zu sagen haben, die erklären, wie Individuen aus strukturellen Möglichkeiten auswählen und damit letztlich zur Variabilität im Sinne von Sapir (1921) beitragen.

Noch weiter gehen gebrauchsbasierte Ansätze, die keine klare Trennlinie zwischen sprachlichem Wissen und Sprachgebrauch sehen. Sie akzeptieren nur tatsächlich im Diskurs belegte sprachliche Beispiele, schließen also elizitierte Daten aus. Als Beispiel seien Thompson & Hopper (2001: 40) genannt, die die vollständige Argumentstruktur, also die strukturell festgelegte Relation zwischen einem Prädikat und seinen Argumenten, als Artefakt bezeichnen, welches “on fabricated examples rather than on corpora of ordinary everyday talk” beruhe.

Die Funktionale Sprachtypologie schließlich gewinnt ihre Abstraktionen oft aus einem größeren Sample von Grammatiken zu einzelnen Sprachen, deren Daten in aller Regel auf einer Kombination von Textkorpora und Elizitation beruhen. Zudem bringen Unterschiede in der Methodik, Analysetiefe und Qualität der einzelnen Grammatiken das Problem sehr stark divergierender Daten mit sich, die es zu integrieren gilt. Dabei

drängen sich häufig Vereinfachungen oder die Ausklammerung bestimmter Daten auf. Diese Datenlage und der Umgang mit ihr dürften ihre Folgen für typologische Generalisierungen haben.

Eine Reihe von Problemen zur Integration von rigider Grammatikalität und Diskurs und zur verfeinerten Darstellung von Grammatikalitätsurteilen scheinen sich mit OT-basierten Ansätzen lösen zu lassen – so zumindest sehen es die Vertreter dieses Ansatzes. In diesem Rahmen schlagen Sorace & Keller (2005) ein Modell vor, welches von durch experimentelle Elizitation gewonnenen Daten ausgeht und unterschiedliche Grade der Grammatikalität in der Syntax integriert. Auf einer solchen Datenbasis lassen sich zwei Typen von Beschränkungen operationalisieren: Eine harte Beschränkung liegt vor, wenn eine Verletzung zu starker Inakzeptabilität führt, die nicht durch Kontexteffekte aufgehoben werden kann. Schwache Beschränkungen dagegen können durch den Kontext aufgehoben werden. Dabei gilt in allen Sprachen der gleiche Beschränkungstyp.

Vier Diskussionspunkte:

Aus der obigen Darstellung ergeben sich die folgenden Diskussionspunkte:

- (i) *Die Verteidigung einzelner theoretischer Ansätze gegen bestimmte, hier durchaus bewusst pauschal genannte Vorwürfe mittels einschlägiger Argumente.*
- (ii) *Die Art der Daten unterscheiden sich je nach Theorie. Welches sind die für bestimmte theoretische Zugänge relevanten Daten und wie gewinnt man sie? In welcher Weise muss man von einem Wechselverhältnis zwischen Daten und Theorie sprechen?*

Die bisherigen Erläuterungen bezogen sich hauptsächlich auf die Formseite (Syntax, Morphologie), nicht erwähnt wurde die Bedeutungsseite:

- (iii) *Welche Probleme der Datenerhebung ergeben sich in der Semantik und in der Pragmatik? Sind die Probleme identisch oder gibt es Unterschiede? Wie sieht es mit etwaigen Wechselverhältnissen zwischen Daten und Theorie aus?*

Wenn unterschiedliche Datenausschnitte unterschiedliche Theorien liefern, stellt sich die Frage, wie eigentlich Grammatikalität beurteilt werden kann. Nach Schütze (1996) ist diese Frage außerhalb rein linguistischer Daten zu lösen:

In principle, there might someday be an operational criterion for grammaticality, but it would have to be based on direct study of the brain, not on human behavior, if it turns out to be possible to discern properties of the mind (e.g., the precise features of grammar) from physical properties of the brain. (Schütze 1996: 26)

Wenn letztlich gerade die Diskussion der Datenerhebung dazu führt, die Frage der Grammatikalität außerhalb rein linguistischer Daten zu beantworten, ergibt sich der folgende vierte Diskussionspunkt:

(iv) *Welche Daten müssen zusätzlich zu den eigentlichen linguistischen Daten zukünftig stärker ins Zentrum des sprachwissenschaftlichen Interesses gerückt werden?*

Literatur

- Belletti, Adriana & Luigi Rizzi (1988). Psych-Verbs and T-Theory. *Natural Language and Linguistic Theory* 6: 291–351.
- Featherston, Sam (2007). Data in Generative Grammar: The Stick and the Carrot. *Theoretical Linguistics* 33: 269–318.
- Manning, Christopher D. (2003). Probabilistic Syntax. In *Probabilistic Linguistics*, Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), 289–341. Cambridge, MA: The MIT Press.
- Newmeyer, Frederick (2003). Grammar is Grammar and Usage is Usage. *Language* 79: 682–707.
- Sapir, Edward (1921). *Language: An Introduction to the Study of Speech*. New York: Harcourt Brace.
- Schütze, Carson T. (1996). *The Empirical Base of Linguistics*. Chicago: University of Chicago Press.
- Sorace, Antonella & Frank Keller (2005). Gradience in Linguistic Data. *Lingua* 115: 1497–1524.
- Thompson, Sandra A. & Paul J. Hopper (2001). Transitivity, Clause Structure, and Argument Structure: Evidence from Conversation. In *Frequency and the Emergence of Linguistic Structure*, Joan Bybee & Paul J. Hopper (eds.), 27–60. Amsterdam: Benjamins.