

Einleitung : Was ist linguistische Evidenz?

Claudia Maria Riehl & Astrid Rothe
Institut für deutsche Sprache und Literatur I, Köln

In der empirischen Linguistik, etwa in der Sprachtypologie und Mehrsprachigkeitsforschung, wird mit verschiedenen Typen von Evidenz gearbeitet. Jeder Typus von Evidenz birgt Vor- und Nachteile, so dass die Aussagekraft der erhobenen Daten in Frage gestellt werden kann. Die Auswahl eines Datentyps als primäre Ressource hängt von den Bedürfnissen der jeweiligen Forschungsrichtung, aber auch von extralinguistischen Faktoren, etwa der Verfügbarkeit von Informanten, ab. Folgende Fragestellungen sind bei der Beurteilung linguistischer Evidenz zentral:

1. Abhängigkeit der Daten von Theorie und Forschungsziel

Nicht jeder Datentyp kann für alle theoretischen Fragestellungen genutzt werden. Untersuchungen zu grammatischen Strukturen einer Einzelsprache verlangen eine andere Datenbasis als diskursanalytische oder textlinguistische Ansätze oder gar Ansätze, die Sprachverarbeitungs- und Sprachproduktionsprozesse untersuchen wollen.

2. Elizitierte vs. natürliche Daten

Ein Corpus mit natürlichen Daten birgt die Nachteile von unkontrollierten Variationen und neben dem Fehlen von negativer Evidenz auch das Problem der statistischen Relevanz. Daten, die im Feld erhoben werden, sind abhängig von der Verfügbarkeit von Informanten, sowie ethischen und kulturellen Besonderheiten der zu untersuchenden Sprachgemeinschaft, was einerseits die Kontrollierbarkeit und andererseits auch den Umfang erheblich einschränken kann. Experimentelles Vorgehen dagegen ist immer mit einem gewissen Grad an Kontrolle über die erhobenen Daten verbunden. Je größer allerdings die Kontrolle, desto größer ist die Gefahr, dass die Daten konstruiert ("unnatürlich") wirken (vgl. den sog. "Lab speech") und somit nicht den realen Sprachgebrauch abbilden. Außerdem grenzen bestimmte Experimente die Wahl von Teilnehmern automatisch ein (z.B. wenn eine Sprache ohne Schriftkultur oder Kindersprache untersucht wird).

3. Positive oder negative Evidenz

Corpora liefern sog. "natürliche" Daten, aber keine negative Evidenz. Diese ist jedoch entscheidend, etwa für die Konstruktion von Kompe-

tenzmodellen. Negative Evidenz kann nicht aus niedrigen Vorkommensfrequenzen geschlossen werden: Corpora können irreführende oder fehlerhafte Verwendungen aufweisen, die häufiger vorkommen, als richtige, aber seltenere Konstruktionen. Die Methode der Introspektion, vor allem mit konstruierten Beispielen, ist problematisch, da Subjekt und Objekt nicht klar unterschieden werden. Dem kann entgegen gewirkt werden, wenn Beispiele mehreren Informanten zur Beurteilung vorgelegt werden. Dabei spielt die Präsentation der zu beurteilenden Daten eine wichtige Rolle. Dies ist vor allem bei gesprochenen Beispielen zu beachten, da die Prosodie unbeabsichtigte Effekte hervorrufen kann. Ein bekanntes Problem sind darüber hinaus individuelle Varianten bei der Akzeptabilität von konstruierten Beispielen. Um dieser Problematik zu entgehen und Varianten zu erfassen, bietet sich eine Abstufung von Grammatikalitätsurteilen an.

4. Linguistische Interpretation von Daten

Nicht nur Umfang und Typus der gesammelten Daten werden durch das Forschungsziel und die zugrunde liegende Theorie vorgegeben, sondern auch die Aufbereitung der Daten. So bestimmt etwa der Abstraktionsgrad der Annotationen, welche Art von Information aus einem Corpus gezogen werden kann. Je nützlicher ein Corpus wird, desto größer wird die Gefahr, dass die Annotationen theorieabhängig sind. Das kann zum Teil durch sog. "multi-tiered labelling" vermieden werden.

Ziel des Kolloquiums "Was ist linguistische Evidenz?" war es daher, Vertreter verschiedener linguistischer Disziplinen zusammenzubringen und die oben genannten Fragestellungen zu diskutieren: Aus den Vorträgen des Kolloquiums wurden in diesem Band Beiträge zusammengestellt, die sich mit zentralen Facetten der Evidenzdiskussion befassen.

Im ersten Beitrag von **Leila Behrens** geht es um eine grundsätzliche Diskussion der linguistischen Evidenz. Sie zeigt auf, wie stark die Evidenz von der zugrunde liegenden Theorie abhängig ist und kommentiert dies anhand eines Beispiels zur Markierung von Aspektualität. Behrens fordert die Etablierung von methodologischen Standards, die cross-linguistisch und disziplinenübergreifend eingesetzt werden können.

In den beiden folgenden Artikeln werden psycholinguistische Vorgehensweisen vorgestellt und diskutiert, die durch experimentell erworbene Daten Evidenzen für kognitive Prozesse der Sprachverarbeitung liefern: Der Beitrag von **Stefan Baumann** diskutiert die Möglichkeit des Einsatzes von Priming-Experimenten und exemplifiziert diese anhand

der prosodischen Markierung von Information unterschiedlicher Gegebenheitsgrade. **Andrea Weber** stellt die Vorteile von *Eye-Tracking-Experimenten* dar und zeigt davon ausgehend die Evidenz für die Verarbeitung von Verstehensprozessen auf.

In einem weiteren Beitrag wird die Problematik von Daten für die Theoriebildung im Bereich grammatischer Strukturen aufgegriffen. **Sam Featherston** zeigt anhand der Methode der *Magnitude Estimation*, wie man mit Grammatikalitätsurteilen zu einer zuverlässigen Datengrundlage kommen kann: Es werden nicht nur die Urteile einer breiten Zahl von Sprechern eingeholt, sondern darüber hinaus auch abgestufte Urteile (nach dem sog. Thermometerverfahren) erfasst. Featherston erläutert diese Methode und ihre Validität anhand von Urteilen zu *wh*-Phrasen.

Der folgende Artikel von **Jürgen Rolshoven** und **Stephan Schwiebert** beschäftigt sich mit der Problematik von künstlich erstellten Daten mithilfe von Computersimulationen. Im Mittelpunkt steht dabei die Corpusrecherche und Dateninterpretation mit Hilfe von Alignment-Strategien, die ebenfalls Evidenzen liefern können. Dieser Beitrag zeigt auf, wie der Typus von Annotation auch die Dateninterpretation beeinflussen kann.

Der abschließende Aufsatz von **Dany Adone** setzt sich mit der Elizitierung authentischer Daten auseinander. Sie diskutiert besonders die Problematik der Datengewinnung in schriftfernen und dem Explorator fremden Kulturen und erläutert hier die kulturell bedingten Schwierigkeiten bei der Datenelizitation (die zu einer Beschränkung von Corpora führen können). Sie wirft dabei die Problematik des kompetenten Sprechers auf: In vielen multilingualen Gemeinschaften, in denen die autochthone Sprache nur noch eine marginale Rolle spielt (z.B. bei den Aborigines in Australien) gibt es viele sog. *semi-speaker*, die eine hohe Variation von sprachlichen Strukturen in ihren Äußerungen aufweisen. Diese Tatsache erschwert die Beschreibung des grammatischen Systems erheblich.

Mit diesem zweiten Band der Reihe des Zentrums 'Sprachenvielfalt und Mehrsprachigkeit' möchten wir einen Einblick in die aktuelle Forschungsdebatte über linguistische Evidenz geben.

Köln, im August 2008

Claudia M. Riehl
Astrid Rothe