

# Welche Fragen können wir mit herkömmlichen Daten beantworten?

MARTIN HASPELMATH

Ich würde hoffen, dass wir LinguistInnen uns recht schnell auf den Konsens einigen können, dass alle Arten von Daten, die bisher verwendet worden sind, im Prinzip für unsere Wissenschaft relevant sind, und dass die Auswahl der empirischen Methode schlicht von der Fragestellung und den vorhandenen Mitteln abhängt. Wer generell behauptet, dass metasprachliche Akzeptabilitätsurteile in der Syntaxforschung nichts zu suchen hätten, oder umgekehrt, dass die Syntax keine Korpusmethoden braucht, irrt.

Eine kleine Anekdote dazu: Ein Kollege erzählte mir einmal, dass ein generativ arbeitender Bekannter, der aus seinem Umfeld nur die Methode der Elizitierung kannte, nach etlichen Jahren der Erforschung einer Sprache der Pazifikregion eine in Texten nicht selten vorkommende Konstruktion einfach nicht kannte. Die InformantInnen brauchten diese syntaktische Konstruktion für die isolierten Sätze einfach und für die Übersetzung aus der Kontaktsprache nicht, während sie im zusammenhängenden spontanen Diskurs schon nach drei Minuten auftauchen würde. Dass ganze Grammatiken von Sprachen wie Deutsch mit künstlichen Beispielen geschrieben werden können, liegt vor allem daran, dass die AutorInnen (als MuttersprachlerInnen und ständige DeutschnutzerInnen) informell ständig mit spontanen Texten umgehen und daher enormes implizites Korpuswissen besitzen. Für die Erforschung noch wenig bekannter Sprachen dagegen ist die Erstellung von Textkorpora eine unverzichtbare Voraussetzung für die grammatische Arbeit.<sup>1</sup>

Auf der anderen Seite ist Elizitierung genauso wichtig, schon für die Morphologie: Komplette Verbparadigmen sind selbstverständlicher Teil einer Grammatik, aber aus Korpora lassen sie sich schwer gewinnen, weil die verschiedenen Formen ganz verschieden häufig sind. Um Formen wie

---

1. In der Feldforschungsszene ist dies seit Franz Boas' Zeiten Allgemeingut: Man spricht von der „Boasschen Trilogie“: Textsammlung, Wörterbuch, Grammatik. Erst wenn für eine Sprache diese drei vorliegen, kann sie als einigermaßen beschrieben gelten.

*stahlst* oder *kämt* belegt zu finden, braucht man riesige Korpora (die mit dem Internet jetzt aber sogar verfügbar sind). Wiederum kann die Illusion, man könnte auf Elizitierung verzichten, nur dann entstehen, wenn man es vor allem mit Sprachen zu tun hat, die seit Jahrhunderten sehr gut erforscht sind und deren Formeninventar bereits durch die Elizitierungsarbeiten früherer Generationen bekannt ist. Dasselbe gilt für syntaktische Paradigmen, und hier liegt meiner Meinung nach der große Beitrag der generativen Syntax: Durch die systematische Anwendung der (Selbst-)Elizitierungsmethode auf Satzmuster konnten viele neue Erkenntnisse gewonnen werden, die früheren Generationen (die sich für die großen Schriftsprachen auf Korpora beschränkten) noch verschlossen waren.

Selbstverständlich haben beide Methoden der Datenerhebung ihre Beschränkungen und Probleme, aber das ist ja gerade die Herausforderung an die Wissenschaft: aus beschränkten und problematischen Daten Erkenntnisse von größerer Tragweite zu gewinnen. Die Vor- und Nachteile von Elizitierungs- und Korpusmethoden sind weithin diskutiert worden, aber hier möchte ich darauf hinweisen, dass spontansprachliche Korpora und Akzeptabilitätsurteile nicht die einzigen Datentypen sind, die SprachwissenschaftlerInnen benutzen. In Abb. 1 habe ich acht klar unterscheidbare Datentypen aufgelistet und klassifiziert. Natürlich könnte man sie anders gruppieren und andere oder feinkörnigere Unterscheidungen machen, aber worauf es mir hier ankommt, ist, dass die linguistische Methodik vielfältiger ist, als es so manche Diskussion unter SyntaktikerInnen erscheinen lässt. (Überhaupt sollte sich die Linguistik weniger syntaktozentrisch verstehen – dies sage ich, obwohl ich selbst Syntaktiker bin.)

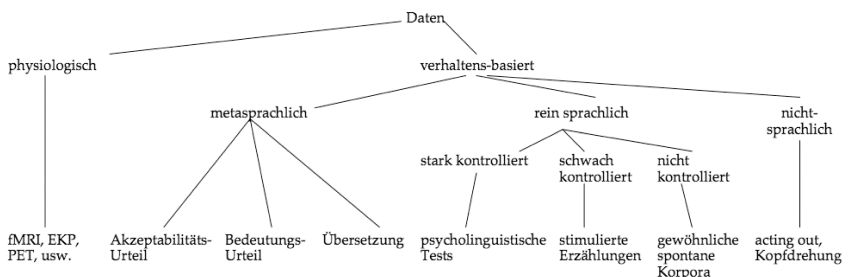


Abbildung 1: *acht Datentypen in der Linguistik*

Die wirklich interessante Frage ist nun, welche Fragen man mithilfe welcher Daten beantworten kann. Von den vielen Fragen, die man über Sprache(n) stellen kann, möchte ich hier nur fünf in Betracht ziehen:

- (1) Welchen Regularitäten gehorcht der Sprachgebrauch einer bestimmten Sprache?
- (2) Welche Regeln enthält die mentale Grammatik der Sprecher einer bestimmten Sprache?
- (3) Was für Arten von Regeln können mentale Grammatiken aus biologischen Gründen überhaupt enthalten?
- (4) Welche Arten von Regularitäten sind in den Sprachen der Welt belegt?
- (5) Warum sind mentale Grammatiken in der Art beschränkt, wie sie beschränkt scheinen (vgl. 3)?
- (6) Warum sind die Regularitäten in den Sprachen der Welt in der Art beschränkt, wie wir es beobachten (vgl. 4)?

Die verhaltensbasierten Daten, mit denen die GrammatikerInnen traditionell arbeiten (also Korpora und Elizitierung), sind ausreichend, um Frage (1) zu beantworten: Wir können Regularitäten des Sprachgebrauchs beobachten und formulieren und wir können anhand neuer Daten bisherige Hypothesen überprüfen. Wir können auch versuchen, durch vergleichende Daten aus aller Welt allgemeine Gesetze über vorkommende und nicht vorkommende Regularitäten aufzustellen (vg. Frage 4), z. B. hinsichtlich der Arten von Lauten, Lautkombinationen, Morphemen und Morphemkombinationen. Die empirische Universalienforschung hat hier schon Vielversprechendes gefunden.

Aber längst nicht so klar ist, ob wir Frage (2) anhand unserer herkömmlichen Daten beantworten können. Wir wissen, dass das Deutsche sowohl in der ersten Person als auch in der dritten Person Plural die Endung *-en* verwendet (*wir komm-en, sie komm-en*), aber welche Regeln haben die SprecherInnen in ihrer mentalen Grammatik? Haben sie zwei homonyme Suffixe *-en<sub>1PL</sub>* und *-en<sub>3PL</sub>*, oder haben sie ein unterspezifiziertes Suffix *-en<sub>PL</sub>*, das aufgrund von allgemeinen Regelanordnungsprinzipien nicht auf die zweite Person Plural angewendet wird? Wir wissen, dass im Englischen sowohl bei Fragesätzen als auch bei Relativsätzen ein Element (*who, which, usw.*) aus seiner normalen Satzposition nach vorne verschoben wird. Aber handelt es sich dabei um dieselbe Regel in unterschiedlichen Kontexten (und deshalb mit etwas unterschiedlichen Auswirkungen), oder handelt es sich um zwei verschiedene Regeln mit ähnlichen Eigenschaften? Solche Fragen werden von GrammatikerInnen, die Frage (2) zu beantworten versuchen, gerne gestellt. Aber sie können nie anhand von herkömmlichen Daten beantwortet werden. Die Argumente sind durchweg „konzeptueller“ Natur, d. h. sie beziehen sich auf die geringere oder größere Ökonomie oder Eleganz einer Beschreibung. Das ist aber weitgehend eine Frage des Geschmacks oder der Forschungstradition, und deshalb scheint sich die Forschung hier oft im

Kreise zu drehen statt voranzuschreiten. Dazu kommt, dass zwei Voraussetzungen für solche Eleganz-Argumente wahrscheinlich gar nicht gegeben sind: Erstens, dass die SprecherInnen selbst beim Spracherwerb die eleganteste Regel erwerben, und zweitens, dass verschiedene SprecherInnen derselben Sprache dieselben mentalen Regeln besitzen. Es kann gut sein, dass mein Deutsch zwei homonyme Suffixe *-en*<sub>1PL</sub> und *-en*<sub>3PL</sub> besitzt, während das Deutsch meiner Schwester ein unterspezifiziertes Suffix *-en*<sub>PL</sub> hat. Welche Art von Daten würde uns helfen, hier weiterzukommen? Ich denke, psycholinguistische Tests (etwa Priming-Experimente) und neurolinguistische Untersuchungen könnten wertvolle Hinweise geben, aber letztlich müsste man in unser Gehirn hineinschauen können, um der mentalen Grammatik auf die Spur zu kommen. Wir bräuchten für eine wirklich mentalistische Linguistik also ganz neuartige Daten, die gegenwärtig nicht oder kaum verfügbar sind.

In den vergangenen Jahrzehnten ist Frage (2) sehr oft in Verbindung mit Frage (3) behandelt worden: Man hat aufgrund von Hypothesen über (2) für bestimmte Antworten auf (3) argumentiert, und umgekehrt. Einzelsprachliche mentale Grammatiken sollten uns Aufschluss über die Universalgrammatik geben, und umgekehrt. Die „Daten“ für die Überprüfung solcher Fragestellungen sind also selbst extrem hypothetisch. Eine universalgrammatische Hypothese, die sich auf einzelsprachliche mentale Grammatiken stützt, müsste eine große Anzahl von verschiedenen Sprachen berücksichtigen, doch wirklich vergleichbare mentalistische Analysen vieler Sprachen gibt es nicht. Das Unternehmen ist extrem aufwendig und oft zirkulär, weil einzelsprachliche Analysen sich typischerweise auf universalgrammatische Hypothesen stützen. Und schließlich wissen wir nicht, ob das Fehlen bestimmter Sprachtypen durch biologische Beschränkungen oder durch einfache funktionale Erwägungen zu erklären ist. Auf diesem Wege ist also kein echter Fortschritt bei der Beantwortung von Frage (3) zu erkennen.

Einfacher wäre es womöglich, zu versuchen Frage (3) mit einem ganz anderen Datentyp beizukommen. Man könnte Versuchspersonen unterschiedliche Arten von sprachlichen Regeln lernen lassen und sehen, welche Regeln sie lernen können und welche nicht. Solche Experimente sind aus praktischen (und vielleicht auch ethischen) Gründen nicht in natürlichen Erstspracherwerbssituationen durchführbar: Wir können uns schlecht neuartige Regeln (oder gar Sprachen) ausdenken und Kleinkinder in ihrer natürlichen Umgebung systematisch mit einem solchen manipulierten Input umgeben. Aber mir scheint, dass künstliche Spracherwerbssituationen (auch mit erwachsenen Sprechern) einen gewissen Aufschluss über biologisch mögliche und nicht mögliche Regelarten geben könnten. Dieser Datentyp scheint jedenfalls für die Beantwortung der Frage (3) (nach der Universalgrammatik) längst noch nicht ausgereizt.

Da wir Frage (3) noch nicht richtig beantworten können, scheint auch der Versuch, Frage (5) zu beantworten, verfrüht (und diejenigen LinguistInnen, die sie im Rahmen des minimalistischen Programms zu beantworten versuchen, sehen diese Gefahr ja durchaus). Und selbst wenn wir Frage (3) nach den biologischen Beschränkungen der Sprachfähigkeit beantworten könnten, hätten wir noch nicht unbedingt genug Daten, um die Warum-Frage (5) zu beantworten. Die Evolutionsbiologie erlaubt es prinzipiell, solche Fragen durch Adaptation zu erklären, aber solche Erklärungen sind vor allem dann überzeugend, wenn es genügend Vergleichsdaten gibt: Immer wieder beobachtet man, dass dem Licht ausgesetzte Organismen Blätter bzw. Augen haben, und deshalb kann man überzeugend argumentieren, dass Blätter und Augen Adaptationen an die Lichteinwirkungen der Umgebung sind. Für die Sprache dagegen fehlen uns schlicht die Vergleichsdaten: Es gibt keine andere Spezies, die über eine vergleichbare biologische Sprachfähigkeit verfügt. Die vergleichende Sprachwissenschaft ist auf den Vergleich innerhalb einer Spezies beschränkt, und höchstens durch spektakuläre Fortschritte in der extraterrestrischen Zoologie wäre hier eine Besserung der Datenlage zu erwarten.

Besser sieht es dagegen mit Frage (6) aus, jedenfalls insofern, als die empirisch beobachteten Beschränkungen nicht durch biologische Beschränkungen bedingt sind (die wir, wie gesagt, schlecht erforschen können). Viele verschiedene menschliche Sprachen haben sich in den vergangenen Jahrhunderten völlig unabhängig voneinander entwickelt, und doch sehen wir eine große Zahl von Ähnlichkeiten, die sich in vielen Fällen durch funktionale Anpassung an die Bedürfnisse der Sprechenden und HörerInnen erklären lassen. Dass Sprachen die overte Akkusativmarkierung z. B. immer wieder auf belebte Objekte beschränken (z. B. *de-n Herr-n*, aber *das-Ø Haus-Ø*), liegt offensichtlich daran, dass die Akkusativmarkierung in diesem Fall für die HörerInnen besonders nützlich ist, weil belebte Objekte eher mit Subjekten verwechselt werden als unbelebte. Diese Erklärung ist vergleichbar mit der Erklärung der Anpassung des Blatts an das Licht, weil wir so viele unabhängige Fälle haben, in denen die Eigenschaft entstanden ist bzw. sich erhalten hat. Gäbe es nur eine einzige Sprache auf der Erde (was ja in ein paar Jahrhunderten durchaus der Fall sein mag), dann hätte so ein Argument nicht viel Überzeugungskraft, weil die Vergleichsdaten fehlen. Um diese funktionale Erklärung mit ganz anderen Daten zu überprüfen, böte sich ein Experiment mit künstlichem Spracherwerb an: Zwei Gruppen von Versuchspersonen bekommen zwei verschiedene künstliche Sprachen zum Lernen vorgesetzt: Eine mit normaler differenzieller Objektmarkierung (Akkusativ nur bei belebten Objekten), und eine mit umgekehrter (Akkusativ nur bei unbelebten Objekten). Die Personen sollten am besten eine

Sprache sprechen, die überhaupt keine Kasusmarkierung hat. Wenn die erste Probandengruppe sich mit dem Lernen signifikant leichter tut, wäre das ein Hinweis darauf, dass die differenzielle Objektmarkierung biologisch vorgegeben ist; wenn nicht, würde das die funktionale Erklärung stützen.

Natürlich fehlen uns für die Adaptations-Erklärung immer noch viele Daten: Idealerweise wären wir in der Lage, die diachrone Entwicklung vieler verschiedener Sprachen im Detail zurückzuverfolgen. Das geht auch in der Biologie nicht richtig, wo die Fossilien-Daten ähnlich spärlich sind wie in der Linguistik. Wir haben auch nicht die Möglichkeit, im Labor künstliche Evolutionsprozesse zu erzeugen. Insofern wird unseren Antworten auf Frage (6) immer etwas Spekulatives anhaften.

Wirklich reich gesegnet sind wir also nur mit Daten für die Fragen (1) und (4), und hier gibt es auch noch sehr viele Lücken: Viele hunderte Sprachen sind noch fast gar nicht beschrieben worden, und tausende nur sehr unvollständig. Selbst gut beschriebene Sprachen wie das Deutsche haben noch keine wirklich umfassenden Grammatiken, die sich auf 10–15 Bände belaufen müssten, um allein das zu erfassen, was im Prinzip der linguistischen Forschung schon bekannt ist. Eine empirisch basierte Linguistik, die sich vor allem an den Fragen (1) und (4) orientiert und die anderen, empirisch schwer zu beantwortenden Fragen nebenbei interessiert zur Kenntnis nimmt, wird sich also nicht langweilen müssen.

*haspelmath@eva.mpg.de*  
*Max-Planck-Institut für evolutionäre Anthropologie, Leipzig*