

Tibor Kiss

Technische Zugänge zum Wesen der Sprache

VOM NUTZEN DER COMPUTERLINGUISTIK
FÜR DIE LINGUISTIK

Einleitung

Sprachwissenschaftler wie Wilhelm von Humboldt (1767–1835), Ferdinand de Saussure (1857–1913) und Noam Chomsky (geb. 1928) betonen, dass zumindest einer der zentralen Gegenstandsbereiche der Sprachwissenschaft die Aufdeckung des menschlichen Sprachvermögens ist – dieses verstanden als ein *Organ*, das es Muttersprachlern gestattet, unabhängig von Bildung und äußeren Bedingungen eine stabile Grammatik zu entwickeln. Diese stabile Grammatik ist eine der Voraussetzungen für das, was wir Sprache nennen. Der Primat der Grammatik führt zu anderen Untersuchungsmethoden und sprachwissenschaftliche Untersuchungen haben daher immer schon einen technischen Charakter: Pioniere der Korpuslinguistik bedienten sich bereits zum Ende des 19. Jahrhunderts mathematischer Modelle zur Bestimmung von Abfolgebedingungen. Allerdings scheiterten diese Untersuchungen zunächst daran, dass einerseits enorme Datenmengen erforderlich waren, um Regelsysteme abzuleiten, andererseits diese Datenmengen aber durch einzelne Forscher nicht mehr bewältigt werden konnten.

Noam Chomsky ersetzte in den 1950er-Jahren den empirischen Zugang durch das Konzept der Introspektion. Dieser introspektive Zugriff – ein *wissen, dass*, nicht jedoch ein *wissen, wie* – erlaubt es Sprechern einer Sprache, zulässige Konstruktionen von unzulässigen abzugrenzen: Sprecher des Deutschen wissen auch ohne jeden Grammatikunterricht, dass der Artikel nicht dem Substantiv folgt, das heißt, dass »Haus ein kaufe ich mir, wenn Lotterie die ich gewinne« kein korrekter deutscher Satz ist. Eine Sprache wird verstanden als die Menge der Sätze dieser Sprache, definiert durch die Grammatik. Hier liegt dann ein mathematisches Modell vor, das sich formal nicht von der Definition einer beliebigen Menge in der Mengenlehre unterscheidet.

Diese Art des Zugangs war außerordentlich erfolgreich – aber auch die Grenzen des Verfahrens sind nun klar erkennbar, sodass wir erneut nach alternativen Zugängen suchen, um das Wirken der Grammatik sichtbar zu machen. Unterstützt durch die Entwicklungen in der Computertechnologie können die enormen Datenmengen, deren Analyse den Forschern des 19. und 20. Jahrhunderts verwehrt geblieben ist, nun bewältigt werden.

Welche Möglichkeiten sich durch einen Einsatz computerlinguistischer Methoden ergeben, soll im Folgenden anhand von zwei Regeln der *Duden*-Grammatik¹ illustriert werden. Zum einen geht es um die Frage, wie die Eigenschaft der Zählbarkeit eines Substantivs zu definieren ist. Zum anderen soll das abstrakte Konzept der Regelmäßigkeit selbst hinterfragt werden.

Die Regel der Zählbarkeit

Die *Duden*-Regel 442 stellt fest, dass »Substantive mit Merkmalkombination »zählbar« plus Singular [...] grundsätzlich immer ein Artikelwort bei sich [haben], und wenn es als letzte Möglichkeit der indefinite Artikel ist.«² In den folgenden Beispielen sind Sätze, die der Grammatik nicht entsprechen, gemäß den Konventionen der Sprachwissenschaft mit einem Stern markiert.

(1a) Ich kaufte gestern ein Buch.

(1b) *Ich kaufte gestern Buch.

(1c) Ich kaufte gestern Bücher.

(1d) Ich kaufte gestern Gold.

(1e) *Ich kaufte gestern ein Gold.

Dass *Buch* ein zählbares Substantiv ist, ist auch dann plausibel, wenn man einen alltäglichen Zugang zum Konzept der Zählbarkeit wählt. In (1a) und (1b) wird es im Singular realisiert. Wenn es dann, wie in (1b), kein Artikelwort bei sich trägt, dann ist das Beispiel ungrammatisch. Ein zählbares Substantiv im Plural, so wie *Bücher* in (1c), bedarf der Realisation eines Artikelworts nicht. Das Substantiv *Gold* ist nicht zählbar, deswegen benötigt es nicht nur keinen Artikel (1d), es kann auch nicht mit bestimmten Artikeln kombiniert werden (1e).

Die *Duden*-Regel 442 ist eine sehr wichtige Regel: Sie ist die Instanz einer universellen Tendenz, nach der zwar Substantive im Plural häufig ohne Artikel erscheinen können, Substantive im Singular aber, wenn sie zählbar sind, ohne Artikel eine deutliche Ausnahme bilden. Vor diesem Hintergrund möchten wir natürlich auch gerne angeben, was *Zählbarkeit* bedeutet.

Definitionsversuche von Zählbarkeit

Hier wäre zunächst die Antwort denkbar, dass es sich um eine Eigenschaft der durch die Sprache bezeichneten Dinge handelt – eine Idee, die der englische Linguist Otto Jespersen in den 1920er-Jahren formulierte. Ein Substantiv wäre dann zählbar, wenn sich das durch das Substantiv bezeichnete Objekt ohne weitere Abgrenzung zu anderen Objekten, die ebenfalls durch das Substantiv bezeichnet werden, zählen lässt. Nach dieser Charakterisierung wäre das Substantiv *Buch* deswegen zählbar, weil man einzelne Bücher voneinander abgrenzen kann und entsprechend dann jedes einzeln abgegrenzte Buch zählen kann. Das Substantiv *Schmuck* wäre aus demselben Grund nicht zählbar, weil man die Abgrenzung nicht durchführen kann. Will

man eine Abgrenzung ausdrücken, muss man sogenannte Numerativkonstruktionen verwenden, in unserem Fall also etwa *zwei Schatullen Schmuck*. Entsprechend sind nicht nur *Bücher* zählbar, sondern auch *Ringe* und *Äpfel*, während hingegen *Schmuck* und *Vieh*, *Gold* und *Wein* aus denselben Gründen nicht zählbar sind.

Leider ist diese Vermutung falsch. Man kann zunächst einmal beobachten, dass Objekte, die in der einen Sprache zählbar sind, dies in einer anderen Sprache nicht sein müssen:

- (2a) Deutsch: *ein Blitz, zwei Blitze*
 (2b) Französisch: *un éclair, deux éclairs*
 (2c) Englisch: **a lightning, *two lightnings, two flashes of lightning*

Und man kann auch in einer Sprache auf dieselben Entitäten mit zählbaren und nicht-zählbaren Substantiven verweisen, etwa

- (3a) *Geschmeide vs. Ringe*
 (3b) *Fußbekleidung vs. Schuhe*

Der in (2) bis (3) beobachtete Kontrast wäre merkwürdig, wenn es eine Eigenschaft der Objekte wäre, um die es hier ginge – dann müsste man ja annehmen, dass Blitze über England physikalisch verschieden sind von Blitzen über Frankreich. Auch wird die Trennung zwischen zählbaren und nicht-zählbaren Substantiven bei Bedarf aufgegeben. Es gibt Regeln, die aus einem zählbaren Substantiv ein nicht zählbares Substantiv machen:

- (4a) Geben Sie *drei Äpfel* in den Salat.
 (4b) Geben Sie *Apfel* in den Salat.
 (5) Millionen Männer bewundern Joaquin Cortez, Tänzer und Freund von Naomi Campbell. Er aber sei zu *viel Mann* für nur eine Frau, soll er gesagt haben.

Linguisten beschreiben die Verwendungsweisen in (4) und (5) durch eine Operation, die sie als *universellen Fleischwolf* beschreiben: Man steckt etwas Zählbares hinein und bekommt die Masse zurück. Die Beispiele verdeutlichen aber auch, dass das Konzept *Wort* möglicherweise nicht der richtige Ort für die Bestimmung des Konzepts Zählbarkeit ist, denn hier geht es ja nicht mehr um Worte per se, sondern um Worte mit einer bestimmten Interpretation!

Und die Interpretation des Wortes hat wiederum einen Einfluss auf die Zählbarkeit, wie anhand des mehrdeutigen Worts *Blech* erläutert werden kann: Das Wort *Blech* bezeichnet ein konkretes Objekt, ein flaches Walzwerksprodukt aus beliebigen Metallen. Aber nicht nur das:

- (6a) Mit dieser Aktion würde sich der schönste Platz von Hohenems zukünftig als ein mit *Blech* zugedachter Platz präsentieren. (→ *Automobil*)
 (6b) Bronze ging an Bertram Fetz, Wolfram Waibel jun. musste trotz zweitbestem Finalergebnis mit *Blech* vorlieb nehmen. (→ *Medaille*)

Interessant daran ist, dass *Blech* in der Lesart *Walzwerksprodukt* pluralisiert werden kann, nicht jedoch in den Lesarten *Automobil* oder *Medaille*, wie die folgenden Beispiele zeigen:

- (7a) Mit dieser Aktion würde sich der schönste Platz von Hohenems zukünftig als ein mit *Blechen* zugedachter Platz präsentieren.
 (7b) Bronze ging an Bertram Fetz, die anderen Sportler mussten trotz zweitbestem Finalergebnis mit *Blechen* vorlieb nehmen.

Die Beispiele in (7) verlangen dem Leser in der Interpretation einiges ab: So kann (7a) nur so verstanden werden, dass der betreffende Parkplatz mit Walzwerkprodukten belegt wird, was aber wieder mit dem Verb *zugedacht* nicht zusammengeht. In (7b) lautet die plausibelste Interpretation, dass den anderen Sportlern als Trostpreise Walzwerkprodukte übergeben wurden. Die Pluralisierung ist in beiden Fällen mit der geforderten Interpretation der Substantive nicht vereinbar.

Zählbarkeitsklassen zur Zuordnung von Substantiven

Wo wäre nun der Ort, um unser Wissen über die Zählbarkeit von Substantiven zu repräsentieren? Zunächst einmal gehen wir davon aus, dass Zählbarkeit keine binäre Eigenschaft ist, sondern, abhängig von der jeweils betrachteten Sprache und gegebenen Kontexten, eine Zahl von Unterklassen angesetzt werden sollte. Für das Deutsche ergibt sich eine Grobklassifikation in drei Gruppen: zählbare Substantive, nicht zählbare Substantive und solche, die nur im Plural realisiert werden können:

- (8a) zählbare Substantive (*Tisch, Lamm, Auto*)
 (8b) nur im Plural realisierte Substantive (*Eltern, Mäse, Finanzen*)
 (8c) nicht zählbare Substantive (*Zorn, Gold, Quatsch*)

Semantische Operationen wie der *universelle Fleischwolf* ermöglichen einen Klassenwechsel, und unterschiedliche Lesarten eines Substantivs führen zu einer Zuordnung desselben Substantivs zu unterschiedlichen Klassen. All dies führt zu der Annahme, dass zumindest eine rein lexikalische Spezifikation der Zählbarkeit nicht sinnvoll ist, wohl aber eine Angabe der Präferenz bzw. der auf eine Lesart eingeschränkten Präferenz. Aber woher wissen wir nun, ob ein Substantiv präferiert zählbar ist?

Analyse der Zählbarkeit in angereicherten Korpora

Die Antwort darauf lautet, dass ein Substantiv dann präferiert zählbar ist, wenn es bestimmte Eigenschaften besitzt, die jeweils kontextuell bestimmt werden können. Als Ausgangspunkt wählen wir hierzu das Vorkommen der betreffenden Substantive in umfangreichen Korpora. Auf der Basis einer Vorverarbeitung extrahieren wir eine Menge von Merkmalen (morphologische und syntaktische Eigenschaften). Diese Merkmale werden dann jeweils auf Zählbarkeitsklassen abgebildet und es erfolgt eine entsprechende Zuordnung der Substantive zu einer Klasse. Das Verfahren ist in Abb. 1 skizziert.

Wie so häufig steckt auch hier der Teufel im Detail. Damit wir überhaupt Merkmale herausfiltern können, sind die folgenden Vorverarbeitungsschritte erforderlich:

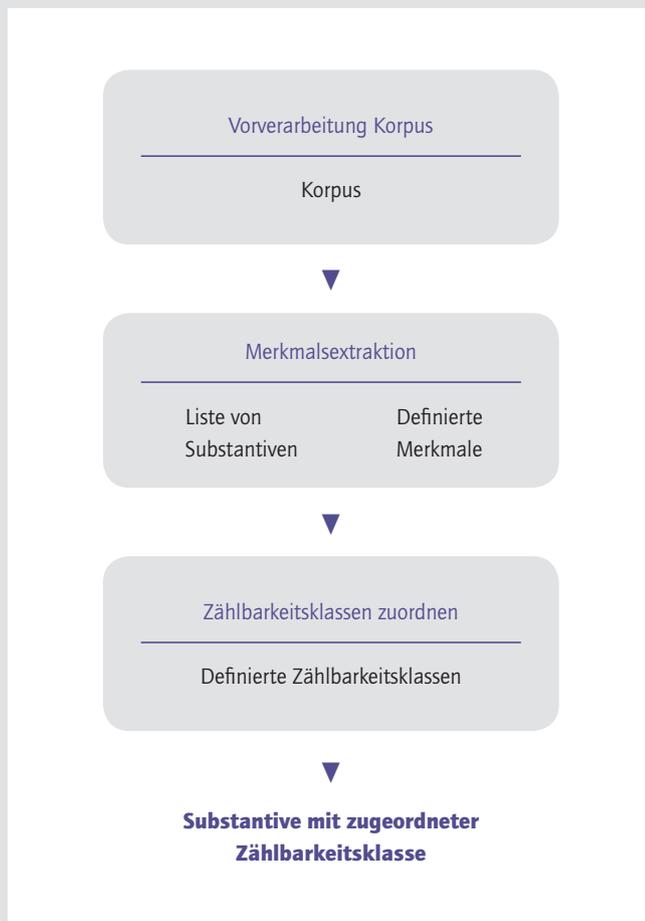


Abb. 1 Zuordnung von Substantiven zu Zählbarkeitsklassen

(9) Wortidentifikation → Satzgrenzenidentifikation → kategoriale Annotation → Kasus-Annotation → flexionsmorphologische Analyse → Lemmatisierung → Numerusannotation

Welche Merkmale sind nun relevant? Hier nur ein Ausschnitt:

- Kongruenz mit einem Verb im Singular *und* Plural (Morphologie)
- Realisation mit indefiniten Artikeln und Numeralen (Syntax)
- Realisation mit Plural wählenden Artikeln (Syntax)
- Realisation mit Artikeln wie *aller, mehr, viel* und Singular (Morphologie, Syntax)

Wie geht man nun mit diesen Eigenschaften um? Hierzu wählen wir ein statistisches Modell, bei dem jede der oben genannten Eigenschaften numerisch abgebildet wird, und zwar, indem man das Vorkommen einer Eigenschaft zählt und ins Verhältnis zu einigen fixierten Größen setzt, etwa dem Vorkommen des Substantivs überhaupt.

Für jedes Substantiv ergibt sich dann eine Repräsentation, in der jeder Eigenschaft ein bestimmter numerischer Wert zwischen 0 und 1 zugeordnet wird, wobei die Werte 1 und 0, die *besitzt diese Eigenschaft* bzw. *besitzt diese Eigenschaft nicht* bedeuten würden, eigentlich nie vorkommen. In den Analysen setzt man etwa hundert unterschiedliche Eigenschaften an, sodass

jedes Substantiv durch einen Vektor in einem Raum mit hundert Dimensionen repräsentiert werden kann. Um dieses Verfahren anzuwenden, ist es zunächst einmal notwendig, eine bestimmte Anzahl von Substantiven (etwa 10 000) nicht nur mit den Merkmalen zu belegen, sondern auch mit der Klasse, die aus diesen Merkmalen folgt. Dies nennt man *Training*. Ist das Verfahren trainiert, reicht es für neue Instanzen, die Merkmale zu bestimmen und man kann sich dann fragen, welcher der Klassen das Substantiv mit den genannten Eigenschaften zugewiesen wird. Abb. 2 verdeutlicht dies anhand von *zwei* Dimensionen:

In der Abbildung soll die Frage beantwortet werden, ob der Stern in der Mitte zur Klasse der blauen Kreise oder zur Klasse der roten Dreiecke gehört. Dazu betrachten wir die durch das weiße Viereck markierte Region um den Stern und zählen hier aus, welche seiner Nachbarn die Mehrheit ausmacht. Unter den $k \geq 15$ Nachbarn besitzen die blauen Kreise die Mehrheit: der Stern wird der Klasse der Kreise zugeschlagen.

Die Illustration zeigt natürlich auch, dass unsere Zuordnung, die auf nur zwei Eigenschaften basiert, fragwürdig ist, es gibt ja nur eine sehr knappe Mehrheit für die eine Klasse. Berücksichtigt man, dass wir anstelle von zwei Eigenschaften etwa hundert Eigenschaften betrachten, wird die Zuordnung deutlich plausibler, verliert aber leider an Anschaulichkeit.

Das hier beschriebene Verfahren ist insbesondere nützlich, um die oben angesprochene Mehrdeutigkeit sichtbar zu machen, denn man kann in der Trainingsphase solche Wörter, die mehr als einer Zählbarkeitsklasse zugehören, auch mehreren Klassen zuordnen und dann die einzelnen Klassifikatoren in einer Sequenz ablaufen lassen. Ein mehrdeutiges Substantiv wird dann auch mehr als einer Zählbarkeitsklasse zugeordnet.

Das Resultat einer solchen Analyse wäre dann nicht nur für die Linguistik, sondern auch für die Sprachgemeinschaft insgesamt sehr nützlich, weil dem Lerner einer Sprache natürlich nicht die Eigenschaften selbst vorgelegt werden können, wohl aber aus den Eigenschaften detaillierte Wörterbücher mit Beispielen abgeleitet werden können, die es einem Lerner des Deutschen dann ermöglichen würden zu verstehen, dass man mit Blechen keine Medaillen meinen kann, mit Blech aber schon.

Das Konzept der Regelmäßigkeit

Ausgangslage

Die zweite Regel aus der *Duden*-Grammatik, die diskutiert werden soll, formuliert nun gerade eine Ausnahme zur Regel 442. Sie lautet wie folgt:

(10) »Viele Präpositionen werden in bestimmten Verbindungen ohne Artikel gebraucht [*an Bord, auf See, außer Haus, außer Konkurrenz, bei Tisch/Hof, binnen Jahresfrist, gegen Morgen, in Not, mit Verlaub, nach Ablauf dieser Frist, nach Wunsch, ohne Gewähr/Garantie/ Zweifel, über Nacht, um Haaresbreite, unter Wasser, von Herzen, vor Anker, zu Bett*], zumal in der Sprache des Verkehrswesens, des Handels und der Verwaltung [*ab Hauptbahnhof, bis Steingasse, auf Bahnsteig 10, ab 10. Klasse, auf Seite 9, auf Bewährung, bei Strafe, bis Waldstadion, gen Norden, in Paragraf 15, infolge starken Nebels, Zimmer mit Bad, laut dpa, nach Durchsicht/Prüfung der Akten, per Luftpost/Nachnahme, pro Kilometer/Stunde, von Amts wegen, zu gegebener Zeit, zwecks Überprüfung*] [...]«³

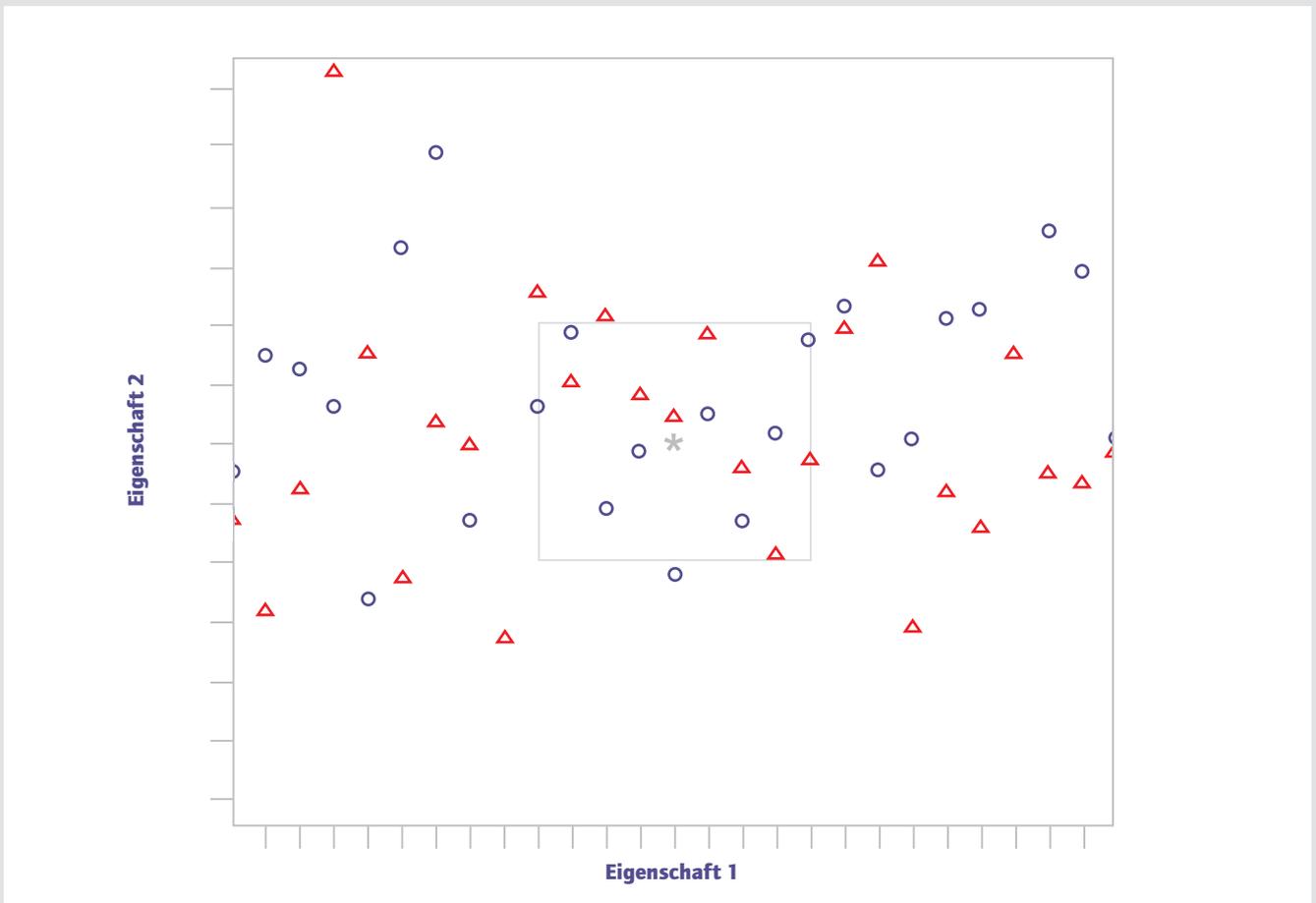


Abb. 2 k-nächste Nachbarn (k = 15)

Nimmt man diese Regel für bare Münze, dann bedeutet sie, dass auch zählbare Substantive im Singular ohne einen Artikel realisiert werden können, diese Formen aber fest sind und nicht regelhaft. Nicht regelhaft bedeutet hier: Man sollte sie aufzählen können. Nun betreten wir mit dieser Fragestellung allerdings eine Grauzone der deutschen Grammatik. Wir wissen nicht genau: Wo endet die Sprache des Verkehrswesens, des Handels oder der Verwaltung und wo fangen die folgenden Verbindungen an:

- (11) auf Aufforderung, durch Beobachtung, in Anspielung,
mit Vorbehalt, ohne Probe, unter Androhung, über Streik
- (12) auf diskrete Aufforderung, durch kritische Beobachtung, in
untertreibender Anspielung, mit leisem Vorbehalt, ohne
positive Probe, unter sanfter Androhung, über regulären Streik

Den Objekten der Präpositionen in (11) und (12) fehlen die Artikel – zur Sprache des Verkehrswesens usw. gehören die Phrasen aber nicht. Auch semantisch sind die Konstruktionen in (11) und (12) transparent, zudem treten die betreffenden Konstruktionen nicht nur, wie (12) zeigt, mit Adjektiven auf, es gibt auch Alternationen zu ganz regelhaften Bildungen mit pluralischen Substantiven, wie die Beispiele in (13) und (14) zeigen.

- (13a) Die im Untergrund operierenden Widerstandsorganisationen beschleunigten das Ende der schwarzen Selbstverwaltungsorgane, indem sie zum Boykott der Lokalwahlen aufriefen und die gewählten schwarzen Gemeinderatsmitglieder *unter Androhung* physischer Gewalt zum Rücktritt zwangen.
- (13b) Milosevic unterschrieb auch *unter Androhungen* von Nato-Bombardementen nicht.
- (14a) Die Arbeitsgemeinschaft Berliner Mieterberatungsgesellschaften [...] hatte bereits Ende Dezember die betroffenen MieterInnen aufgefordert, die Mieterhöhung im Januar nur *unter Vorbehalt* zu zahlen.
- (14b) Zudem wird der Kaufvertrag *unter Vorbehalten* abgeschlossen: So müssen die Baubewilligung sowie ein Verkehrsgutachten und rechtsgültige Strassenprojekte zur Lösung der Verkehrsprobleme im Umfeld des Stadions vorhanden sein.

Die Überprüfung der Regelhaftigkeit einer Konstruktion kann nur experimentell oder im Rahmen einer empirischen Analyse umfangreicher Daten erfolgen.

Feststellung der Regelmäßigkeit

Das Verfahren, das hier knapp vorgestellt werden soll, macht recht komplexe mathematische Vorannahmen. Die Grundidee ist jedoch relativ einfach. Wenn eine bestimmte Kombination regelhaft ist, dann kann sie prinzipiell unendlich viele Instanzen haben, was wiederum bedeutet, dass die Zahl von Kombinationen, die in einem umfangreichen Text nur einmal vorkommt – die *Hapax legomena* (Singular: *Hapax legomenon*, altgr. für »einmal gesagt«) –, bei steigender Textgröße nicht sinkt. Wir definieren hierzu ein Produktivitätsmaß $P(N)$, das als Verhältnis der erwarteten *Hapax legomena* zur Gesamtzahl der vorkommenden Konstruktion, das heißt $P(N) = E[V(1, N)]/N$ (E bezeichnet hier den statistischen Erwartungswert).⁴ Die natürliche Interpretation dieses Maßes ist die Wahrscheinlichkeit, mit der man einen neuen Ausdruck einer Klasse findet, nachdem man schon N Worte betrachtet hat.

Wenn eine Kombination nicht regelhaft ist, wird man irgendwann alle Instanzen dieser Kombination einmal gesehen haben und irgendwann dann auch mehr als einmal, was wiederum bedeutet, dass $P(N)$ irgendwann unter einen Schwellwert fällt und diesen auch nicht wieder überschreitet. Der negative Fall ist leichter zu erklären und auch zu visualisieren als der positive und soll anhand des Würfels erläutert werden. Eine Kombination kann dann als regelhaft bewertet werden, wenn die folgenden Bedingungen erfüllt sind:

- Das Produktivitätsmaß $P(N)$ fällt nicht unter einen Schwellwert.
- Das Vokabular steigt mit steigendem Korpus.
- Die einmal vorkommenden Elemente des Vokabulars steigen mit steigendem Korpus.

Die Überprüfung erfolgt erneut durch die Analyse umfangreicher Korpora, wobei diese so weit aufbereitet sind, dass Instanzen zählbarer Substantive zumindest annähernd identifiziert wurden. Im vorliegenden Fall werden Daten aus der *Neuen Zürcher Zeitung* und der *Frankfurter Rundschau* verwendet; insgesamt handelt es sich um etwa 200 Millionen Wörter. Um die genannten Werte zu bestimmen, besteht die Möglichkeit einer empirischen Bestimmung, wobei für gegebene Teilkorpora jeweils die oben genannten Kenngrößen bestimmt und ins Verhältnis zueinander gesetzt werden und an die Stelle von $E[V(1, N)]$ die empirischen Werte für $V(1, N)$ gesetzt werden und andererseits die mathematische Bestimmung des Produktivitätsmaßes $P(N)$ auf der Basis einer Schätzung für $E[V(1, N)]$.

Ausblick und Perspektive

Betrachtet man die Ergebnisse, muss man feststellen, dass das Bild der deutschen Grammatik, das an dieser Stelle im *Duden* skizziert wird, etwas zu einfach ist. Offensichtlich gibt es regelhafte Konstruktionen, die der *Duden*-Regel 442 zuwiderlaufen, die andererseits aber eben wegen ihrer Regelmäßigkeit nicht durch den Versuch einer Aufzählung – so wie dies in der Regel 395 geschieht – erfasst werden können. Vorläufig können wir auf der Basis annotierter und aufbereiteter Korpora und der computerlinguistischen Verfahren nur dieses Problem feststellen, ohne angeben zu können, welches die Bedingungen sind, aufgrund derer eine regelhafte Kombination von Präposition und zählbarem Substantiv im Singular, allerdings ohne entsprechende Artikelwahl, möglich wird. Das Ziel ist es, die determinie-

renden Faktoren für die Kombination von Präposition und Substantiv auf der Basis einer breiten empirischen Untersuchung zu identifizieren. Hierzu muss nahezu der gesamte Bestand der Präpositionen des Deutschen in die Untersuchung einbezogen werden, wir planen dies gerade für Korpora in der Größe von etwa 350–500 Millionen Wörtern.

Diese Korpora werden auf unterschiedlichen Ebenen um Merkmale angereichert (in der Analyse des Wortes, der Satzstruktur und der Satzbedeutung) und auf der Basis der angereicherten Merkmale analysiert. Betrachtet man unter diesen Bedingungen etwa die Verteilung der Präposition *unter*, so kann man feststellen, dass sie außerordentlich viele Lesarten besitzt, eine Vielzahl dieser Lesarten – insbesondere die räumlichen (die man möglicherweise als prototypisch bezeichnen möchte) – in den Kombinationen mit einem Substantiv aber wegfällt, während Lesarten wie »Bedingung« oder »Art und Weise« sehr stark zunehmen.

Im Idealfall werden wir also die relevanten Faktoren aus umfangreichen Daten herausfiltern und entsprechend für die theoretische Linguistik und die Grammatikschreibung bereitstellen. Es ist aber auch sehr gut möglich, dass wir schlussendlich anerkennen müssen, dass es zwar einen prototypischen Kern von Eigenschaften gibt, der die Bildung der genannten Sequenzen steuert, dass aber Neubildungen durchaus auch davon abweichen können, das heißt, dass eine vollständige Erfassung dieses Phänomenbereichs durch symbolische Regeln nicht möglich sein wird und ab einem bestimmten Punkt nur noch Heuristiken formuliert werden können. Dies mag zunächst enttäuschend aussehen, ist aber tatsächlich gegenüber dem aktuellen Stand der Forschung ein deutlicher Fortschritt: Zwischen den *Duden*-Regeln 442 und 395 besteht derzeit eine grammatische Grauzone. Reduzieren wir diese auf ein angemessenes Maß, sind unsere Kenntnisse über das Wesen der Sprache erneut erweitert worden.

Anmerkungen

- 1 Duden Grammatik 2005.
- 2 Duden Grammatik 2005, S. 337.
- 3 Duden Grammatik 2005, S. 306.
- 4 Diese Idee stammt von Harald Baayen, vgl. Baayen 2001.
- 5 Vgl. Kiss 2007.