

bereitet die linguistischen Theorien so auf, dass Computerlinguisten sie je nach Wunsch in konkrete Anwendungen ummünzen können. Gleichzeitig werden die Theorien während des Formalisierungsprozesses so geschärft, dass sich für Linguisten neue Perspektiven, Probleme und Forschungsziele

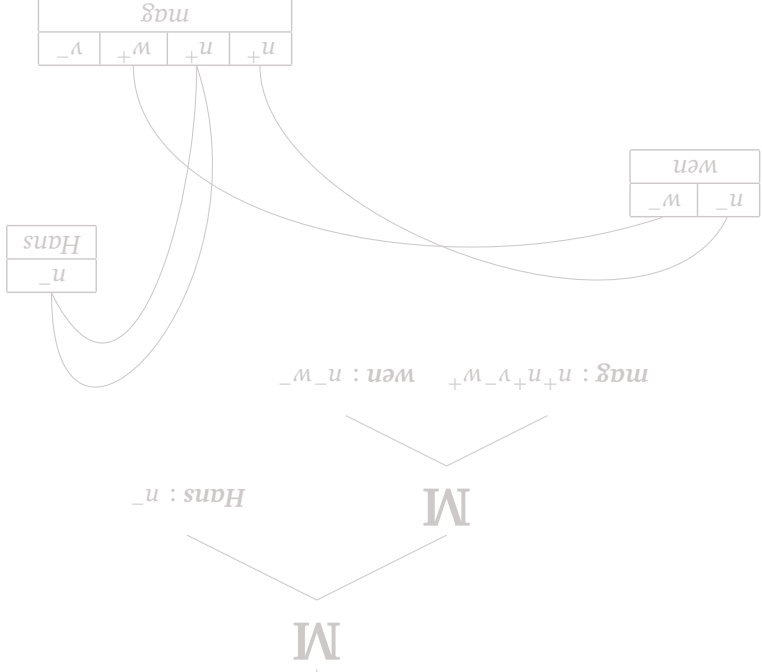
eröffnen.

Die in der mathematischen Linguistik entwickelten Werkzeuge finden mitunter auch Anwendung in anderen Bereichen der Naturwissenschaft. Am spannendsten ist der Einsatz bestimmter Formalismen (den Minimalistischen Grammatiken artverwandt) in der Biologie, wo sie zur Genomsequenzierung und Modellierung von Proteinen und Messengern-RNA herangezogen werden. Doch es bestehen auch Verbindungen zur reinen Informatik (z.B. der Spezifikation und Konvertierung von XML-Varianten).

Der Anwendbarkeit linguistischer Forschung sind im Prinzip keine Grenzen gesetzt. Bedenkt man die zentrale Rolle, die Sprache in unserem Leben spielt, privat wie beruflich, zeigt sich, dass es kaum einen Bereich gibt, der nicht von linguistischer Expertise profitieren könnte.



## Von Mathematischer Linguistik O und anderen obskuren



Optimalitätsbedingungen, weil er an ihnen die Nachvollziehbarkeit der schrittweisen Berechnungen vermisste. Erst später erschlossen sich ihm die Vorteile der Methode.

Viele Linguisten stehen Optimalitätsbedingungen ebenso skeptisch gegenüber, doch ihre Skepsis speist sich aus falschen Vermutungen über ihre Eigenschaften. Ich hoffe, dass meine Resultate diese verhärteten Fronten aufbrechen werden, sodass auch in der Linguistik Optimalitätsbedingungen zukünftig eine größere Rolle spielen werden. Denn eines ist umstritten (weil ich es bewiesen habe): Genau wie ihre Kollegen aus der Physik sind linguistische Optimalitätsbedingungen einfacher als ihre rein strukturellen Gegenstücke.

#### 4 Und wofür ist das gut?

Die mathematische Linguistik hat vielerlei Anwendungsbereiche, zuvorderst natürlich in der Computerlinguistik, also jenem Teil der Linguistik, der sich mit der Entwicklung von Software und Programmen zur Verarbeitung und Erzeugung von Sprache auseinandersetzt.

Computerlinguisten zeichnen verantwortlich für die Text-zu-Sprache Programme für Blinde, automatisch generierte Wetterberichte, Spracherkennungs- und Diktiersoftware wie *Dragon Naturally Speaking*, maschinelle Übersetzung á la *Google translate*, sowie, wenngleich die Vertreter dieser Zunft es nicht gerne zugeben, die Rechtschreibkontrolle von Microsoft Office. Und erst vor wenigen Tagen ist im naturwissenschaftlichen Fachmagazin *Science* ein Artikel erschienen, der das Potential computationeller Analyse großer Textsammlungen für die Geistes- und Sozialwissenschaften aufzeigt. Unter anderem lassen sich heutzutage erstmals die schrittweise Verbreitung von Modetrends und politischen Strömungen in einem historischen Kontext empirisch auswerten, was auch für eher profane Anliegen wie Marketing, Werbung und PR von großer Wichtigkeit ist.

Ohne die mathematische Linguistik wäre die Computerlinguistik in einer ähnlichen Lage wie der Ingenieur ohne den Physiker: Zwar sind Tricks und Kniffe bekannt, die in der Praxis hinreichend funktionieren, doch die Entwicklung neuer Verfahren sowie die Grenzen der Bekannten könnten nur durch Trial-and-Error ermittelt werden. Der Physiker kann diese Probleme auf einer theoretischen Ebene lösen. Dabei macht er mitunter vereinfachende Annahmen, die der Ingenieur unter Einsatz seiner praktischen Erfahrung wieder ausgleichen muss, doch nichts desto minder besteht eine für beide Seiten profitable Symbiose.

Mit ihrem starken Fokus auf reine Theorie gestaltet die mathematische Linguistik die Landschaft der Computerlinguistik entscheidend mit. Sie

zurückgegriffen, wenn ein Wortmolekül, das eigentlich stabil sein sollte, es nicht ist, also der entsprechende Satz von Muttersprachlern für falsch befunden wird. Eine auf Optimalitätsbedingungen basierende Erklärung postuliert, dass Wortmoleküle miteinander im Wettkampf stehen und nur die Besten diesen Wettkampf überleben. Die Stabilität von Wortmolekülen ist somit eine Vorbedingung für die Grammatikalität von Sätzen, aber sie kann sie nicht garantieren.

In meiner Suche nach einem formalen Modell für Optimalitätsbedingungen fiel mir auf, dass die miteinander konkurrierenden Wortmoleküle starke strukturelle Ähnlichkeiten aufweisen. Eine Idee begann in meinem Kopf heranzuwachsen, aus dem sie zu verdängen selbst ein Wochenende trip nach Death Valley nicht vermochte. Nach einem exzentrischen Nachmittag, den ich damit zubrachte, die Sanddünen des Todes als in überdimensionalen Lettern mit den relevanten Beweisen zu verzieren und jeden anzukreuzen, der mein Werk rücksichtslos zu zertrampeln drohte, fand sich mein Verdacht bestätigt: Die strukturellen Ähnlichkeiten sind so stark, dass alle wetteifernden Wortmoleküle sich als eine einzelne abstrakte Struktur darstellen lassen. Dies ist vergleichbar mit dem Schritt von Wortmolekülen zu Aufbauprozeduren. Das Optimalitätskriterium wiederum lässt sich auf dieser abstrahierten Ebene als normale strukturelle Bedingung ausdrücken. Von diesem Punkt an genügt ein Modikum höherer Mathematik um zu zeigen, dass die in der Literatur gängigen Optimalitätsbedingungen eigentlich keine sind. Stattdessen lassen sie sich auch direkt als strukturelle Bedingungen auf Wortmoleküle ausdrücken, zum Beispiel durch das Einführen neuer Merkmale. In gewisser Weise sind Optimalitätsbedingungen also bloß eine eleganter Methode, diese strukturellen Eigenschaften einzufordern. Eine ähnliche Situation findet sich in der Physik. Möchte man beispielsweise den Pfad eines geworfenen Balls vorhersagen, kann man dies auf zwei Arten tun. In der klassischen Mechanik berechnet man für jeden Zeitpunkt separat Energie, Reibung, Schwerkraft, Winkel etc. um so Punkt für Punkt den Pfad abzutragen. In der Lagrange-Mechanik hingegen betrachtet man alle möglichen Pfade und nimmt dann den Pfad mit der geringsten „Wirkung“ (ein physikalischer Fachbegriff). Ebenso kann man den Pfad, den Licht durch ein Medium wählt, schrittweise berechnen, oder man wählt aus allen möglichen Pfaden den kürzesten. In beiden Fällen ist zweite Methode bedeutend einfacher. De facto wären viele Errungenschaften der modernen Physik ohne solche Optimalitätsbedingungen undenkbar, weil schrittweise Berechnungen zu kompliziert wären. Ein beliebtes Beispiel hierfür ist QED, die Theorie der Quantenelektrodynamik, die selbst mit Optimalitätsbedingungen noch ungeheuer komplex ist. Ironischerweise war ihr Erfinder Richard Feynman in seinen College-Jahren kein Freund von

*A good scientific theory should be  
explicitable to a barnard.*  
— Ernest Ruthenford  
1908 Nobelpreisträger für Chemie

*Frohe Weihnachten 2010*

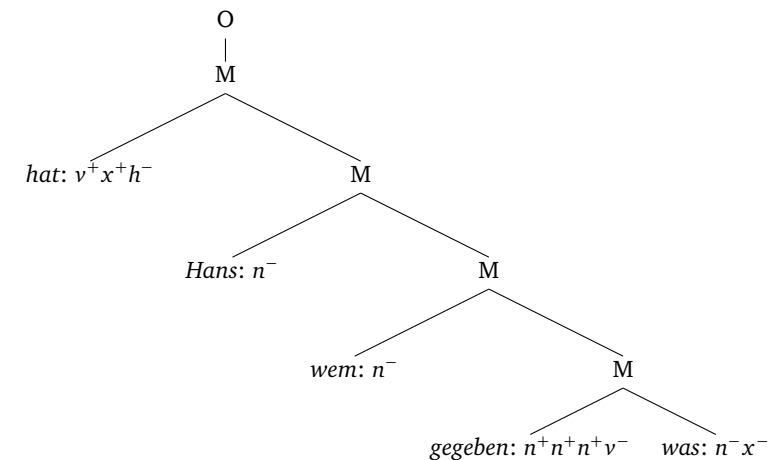


Abbildung 2: Aufbauprozedur für das Wortmolekül in Abb. 1. M signalisiert die Verschmelzung zweier Kategoriemerkmale, O die Verschmelzung zweier Lizenzierungsmerkmale

Aufbauprozeduren haben aber auch Potential in Bezug auf praktischen Anwendungen, Potential, das sich aus ihrer strukturellen Einfachheit im Vergleich zu Wortmolekülen speist. Der Rechenaufwand für die Verarbeitung von Sätzen durch den Computer, zum Beispiel, kann oft um den Faktor 100 oder mehr gesenkt werden, indem nur Aufbauprozeduren berechnet werden anstatt ganzen Wortmolekülen. Und da beide Stellschrauben auf der Ebenen von Aufbauprozeduren identisch sind, kann auch jede Strategie, die auf den letzteren aufsetzt, für beide verwendet werden.



In den letzten paar Seiten haben wir gesehen, dass sich viele interessante mathematische Generalisierungen über Sprache finden lassen. Ob es nun um die Verteilung spezifischer Wörter wie *je* und *durchaus* oder strukturellen Eigenschaften von Sprache im Allgemeinen geht, ohne Mathematik lassen sich keine zufriedenstellende Antworten finden, nein, oft sind selbst die Fragen ohne eine mathematisch geschärfte Theorie nicht formulierbar.

### 3 Und deine Arbeit?

Meine eigene Forschung beschäftigt sich mit dem Status von *Optimalitätsbedingungen* in Sprache. Auf solche Bedingungen wird manchmal

die Essenz schwach kontextsensitiver Systeme: Es gibt keinen Bann gegen sich kreuzende Pfade.

Es konnte gezeigt werden, dass Minimalistische Grammatiken genau die Klasse schwach kontextsensitiver Systeme beschreiben. Minimalistische Grammatiken treffen den Nagel also auf den Kopf, auch wenn der Nagel ein bisschen größer ist als Sprache (weil nicht jedes schwach kontext-sensitive System sich wie menschliche Sprache verhält, zum Beispiel eben eines mit zwölf Lizenzierungsmerkmalen auf einem einzelnen Wort, wie wir weiter oben bereits diskutiert haben). Dass die Mächtigkeit des Formalismus also mit den beobachteten Phänomenen annähernd in Einklang ist, bläst Wind in die Segel der Anhänger Minimalistischer Syntax — die Theorie ist offenbar nicht zu weit entfernt von der (soweit bekannten) Realität.

Dabei darf man jedoch nicht voreilig schließen, dass Sprache sich wirklich um Wortmoleküle und Brücken dreht, wie das folgende Resultat zeigt. Lange Zeit wurde unter Linguisten hitzig debattiert, wie Lizenzierungsmerkmale auf die Struktur einwirken. Eine Schule vertrat — freilich in einer anderen Terminologie — die oben ausgeführte Ansicht, dass bestimmte Stellungsbrücken mitunter gesprengt werden. Die andere meinte, dass jede Stellungsbücke erhalten bleibt. Mathematisch betrachtet ist der zweite Ansatz komplexer als der erste, zumindest wenn man die Struktur der letztendlich geformten Wortmoleküle im Auge hat. Es zeigt sich aber, dass beide Strukturen auf einer abstrakteren Ebene identisch sind: ihrer Aufbau-prozedur. Die Aufbau-prozedur listet einfach, in welcher Reihenfolge sich welche Merkmale miteinander verbunden haben. Ob nun Stellungsbücken abgerissen werden oder nicht ist hierfür unerheblich, denn so oder so haben wir es nur mit der Verschmelzung von zwei Lizenzierungsmerkmalen zu tun. Der Unterschied zwischen den Stellungsbücken-Abreißen und den Stellungsbücken-Erhalten liegt also bloß darin, wie Aufbau-prozeduren in Wortmoleküle umgewandelt werden.

Ein Grund, weshalb der Streit nicht ohne mathematische Hilfsmittel geschlichtet werden konnte, ist der Mangel an unstrittigen Daten, die eines der Stellungsbücken-Szenarien klar widerlegen würden. Nun ist es selbstredlich eine Binsenweisheit, dass eine jede Menge an Daten nur einen kleinen Ausschnitt der Wirklichkeit präsentiert, aus dem Fehlen von Gegenbeispielen also nicht die Richtigkeit einer Hypothese folgt. Wenn aber auch langjähriges Suchen keine relevanten Daten zutage fördert, dann liegt die Vermutung nahe, dass die zur Debatte stehende Unterscheidung letztendlich irrelevant ist. Frank gesprochen: Wenn es keine Evidenz gegen einen der beiden Ansätze gibt, dann liegt das vielleicht daran, dass nicht Wortmoleküle sondern Aufbau-prozeduren die Objekte sind, die den sprachlichen Prozessen im Gehirn zugrunde liegen.

## 1 Was macht ein Linguist?

Wer wissen will, was ein Linguist denn auf die Kosten des Steuerzahlers in seinem Glasturm so treibt (Elfenbein kann sich die Wissenschaftlerge-parenz ja mittlerweile auch ganz groß geschrieben), muss zuerst verstehen, warum der folgende Spruch kein Widerspruch ist: Linguisten studieren Sprache, nicht Sprachen.

*Sprache* bezeichnet hier die Sprachfähigkeit des Menschen, die nicht mit spezifischen Sprachen wie Deutsch, Englisch, Mandarin-Chinesisch, Quechua, Tagalog oder Inkuitut gleichzusetzen ist. Vielmehr sind die Sprachen, die auf unserer Erde gesprochen wurden und werden, ein Produkt dieser Sprachfähigkeit.

Aber was ist denn nun mit Sprachfähigkeit gemeint? Das ist eine der großen ungelösten Fragen der Linguistik. Wir wissen, dass jedes sich normal entwickelnde Kind die Sprache seiner Umgebung spielerisch erwirbt, ganz ohne dass die Eltern ihm Vokabular, Formenlehre oder Satzstellung in stundenlangem Frontalunterricht einpauken müssten. Beispielsweise findet jedes deutschsprachige Kind früher oder später heraus, dass (1a) und (1b) tadellose Sätze sind, während (1c) von Erwachsenen nicht akzeptiert wird (was hier mit einem \* angezeigt wird).

- (1) a. Der Hans hat den Martin gehauen.  
b. Die Maria hat gesagt, dass der Hans den Martin gehauen hat.  
c. \* Die Maria hat gesagt, dass der Hans hat den Martin gehauen.
- Es weiß darüber hinaus noch, dass mit *weil* statt *dass* beide Varianten möglich sind.

- (2) a. Die Maria ist traurig, weil der Hans den Martin gehauen hat.  
b. Die Maria ist traurig, weil der Hans hat den Martin gehauen.
- All das und noch viel mehr weiß ein Kind, ohne dass ein Erwachsener es ihm explizit beigebracht hätte. Wie bemerkenswert das ist, wird klar, wenn man sich bewusst macht, welch einfache Dinge Kinder für gewöhnlich nicht von selbst lernen: Lesen, Zahlen über 10, Multiplikation und Schwimmen, ja leider nicht einmal Tischmanieren. Selbst was die meisten Menschen als allererstes mit Sprache verbinden, das „gute Deutsch“ (Rechtschreibung, Fall-Regeln, dass man nicht *tun* sagt, wie man den Konjunktiv gebrauch-t und dass (2b) oben nicht grammatisch sei), bereitet ihnen Schwierig-keiten. Es scheint, als ob Kinder von Natur aus darauf spezialisiert wären, Sprache zu lernen.

Tatsächlich haben 40 Jahre linguistische Forschung Erstaunliches zutage gebracht. Bereits im Mutterleib können Kinder ihre Muttersprache von anderen anhand der Satzmelodie unterscheiden. Sie wissen von Anfang an sprachliche Laute von nicht-sprachlichen wie beispielsweise Räuspern, Husten oder Lachen zu unterscheiden — die dabei zu umschiffenden Untiefen kennt, wer schon einmal versucht hat, in Kopenhagen eine Pizza zu bestellen (Dänisch, ein eloquentes Gewürge). Später lernen sie bis zu 10 neue Wörter pro Tag, sodass Tafelklassler über einen aktiven Wortschatz von rund 5000 Wörtern verfügen, womit sie in etwa 95% eines durchschnittlichen Zeitungsartikels verstehen können (die fehlenden 5% setzen sich primär aus abstrakten Begriffen, Fremd- und Fachwörtern zusammen). Das Bild, das sich abzeichnet, ist das eines *Sprachinstinkts*: Kinder lernen instinktiv zu sprechen, genauso wie sie ganz natürlich zu laufen lernen.



Auf den ersten Blick mag die Idee eines Sprachinstinkts selbstverständlich erscheinen; allein, in der Natur ist nichts selbstverständlich. Wenn aber erst einmal eine Entdeckung gemacht wurde, so unbemerkenswert sie auch sein mag, führt sie weiter als der naive Betrachter sich auszumalen vermag. In unserem Fall existieren faszinierende Parallelen zwischen der Entwicklung von Sprache und der anderer Instikte.

Der Sehinstinkt, beispielweise, ist bei Geburt stark unterentwickelt. Ein Säugling muss sich in den ersten Wochen nach seiner Geburt mit einem Schwarz-Weiß-Bild zufriedengeben, dann kommen langsam Rot und Gelb hinzu, ehe das volle Farbspektrum wahrgenommen werden kann. Auch Formen und Umrisse sind anfänglich nur unscharf, und die Tiefenwahrnehmung lässt ebenso zu wünschen übrig. Diese graduelle Entwicklung entspricht dem schrittweisen Spracherwerb: von der Lallphase, in der die Artikulationstechniken im Allgemeinen einstudiert werden, zur Ein- und Mehrwortphase, bevor schließlich die Lust, sich mit seinen Eltern in vollständigen Sätzen auszutauschen, vom Kind im Alter von 3 Jahren entdeckt und mit Einsetzen der Pubertät zu Grabe getragen wird.

Doch es existiert noch eine weitere interessante Parallele: ohne Trainingsmaterial wird der Instinkt nicht erworben. Kinder, die von Geburt an in kompletter Dunkelheit gehalten werden (bedauerlicherweise gibt es tatsächlich solche Fälle), entwickeln ihren Sehsinn nicht weiter, nein, er verkümmert sogar, sodass sie erblinden. Dieser Schaden kann später nicht mehr repariert werden, der Sinn muss in einem bestimmten Zeitraum voll ausgebaut werden. Nachdem dieser verstrichen ist, kann das Versäumte nicht mehr nachgeholt werden. Das gleiche trifft auf den Spracherwerb zu: Wenn nie mit dem Kind gesprochen wird, und es auch nicht Gele-

Die wohl wichtigste betrifft die Komplexität der Wortverbindungen, welche schon seit langer Zeit studiert wird. So gibt es empirische Resultate, die Sprache in der sogenannten Klasse *schwach kontext-sensitiver* Systeme verorten. Diese Klasse ist eine konservative Erweiterung der *kontext-freien* Systeme, zu denen die meisten Programmiersprachen und die Sprache arithmetischer Formeln gehören. Kontext hat hier nichts mit der gemeingeläufigen Bedeutung von Kontext zu tun. Ein kontext-freies System kann in beschränkten Maße zählen. Beispielsweise kann es sicher stellen, dass in jeder arithmetischen Formel genauso viele öffnende wie schließende Klammern stehen. Ein einfaches Beispiel gibt es im Deutschen.

- (9) a. Der Peter sagt, dass der Hans läuft.
- b. Der Peter sagt, dass der Paul den Hans laufen lässt.
- c. Der Peter sagt, dass der Josef den Paul den Hans laufen lassen sieht.
- d. Der Peter sagt, dass die Rosa den Josef den Paul den Hans laufen lassen sehen lässt.
- e. Der Peter sagt, dass die Maria die Rosa den Josef den Paul den Hans laufen lassen sehen lassen sieht.

Im Prinzip lassen sich beliebig lange Sätze dieser Art konstruieren, wobei jedoch die wirklich langen so lang wären, dass das Universum in sich zusammenfallen würde ehe der gesamte Satz mit einer dem Menschen verständlichen Sprechgeschwindigkeit von Anfang bis Ende in einem Rutsch vorgetragen werden könnte. Dabei liegt allen Sätzen ein einfaches Muster zu Grunde: Im Gliedsatz finden wir erst eine Sequenz von Nomen, gefolgt von einer Sequenz von genauso vielen Verben. Das letzte Nomen hängt vom ersten Verb ab, das vorletzte vom zweiten, das vorvorletzte vom dritten, und so weiter. Dabei ist es möglich, die Verben und die von ihnen abhängigen Nomen so mit Linien zu verbinden, dass diese Linien sich nie kreuzen (der Leser ist eingeladen, den praktischen Beweis mittels Bleistift anzutreten). Das ist die Essenz kontext-freier Systeme: Die Möglichkeit, die von einander abhängigen Elemente mit Pfaden zu verbinden, die sich nicht kreuzen.

Aber wie gesagt ist Sprache nicht kontext-frei sondern schwach kontext-sensitiv. Heißt das, dass es in der Sprache manchmal Pfade gibt, die sich unausweichlich kreuzen? In der Tat, und das bekannteste Beispiel ist das niederländische Gegenstück zu unseren Beispielsätzen. Denn im Niederländischen stehen die Verben in umgekehrter Reihenfolge. Statt *laufen lassen sehen lassen sieht* haben wir also *sieht lassen sehen lassen laufen*. Das heißt, das letzte Nomen hängt vom letzten Verb ab, nicht vom ersten, das vorletzte vom vorletzten, nicht vom zweiten, und so weiter. In diesem Fall ist es nicht möglich, die Wörter so zu verbinden, dass sich keine Pfade kreuzen. Das ist

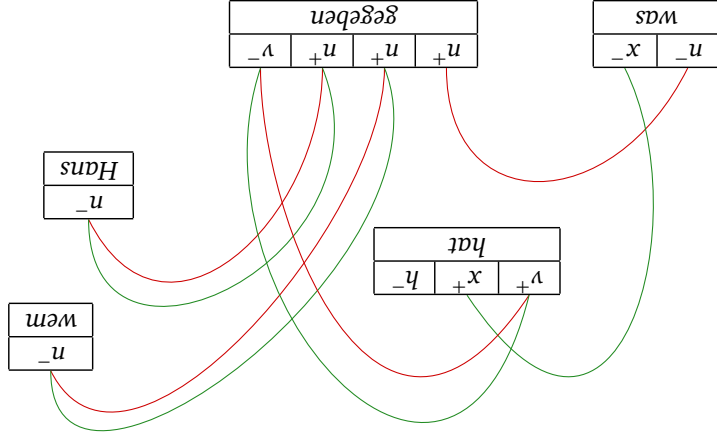


Abbildung 1: Wortmolekül für Was hat Hans wenn gegeben?. Bedeutungsbrücken in rot, Stellungsbrücken in grün

Natürlich verbietet uns nichts in der vorliegenden Theorie, ein entsprechendes Pendant von *hat* mit zwei  $x^+$  der Liste hinzuzufügen. Das ist per se kein Mangel der Theorie; diese Liste von Wörtern wäre bloß keine deutsche mehr, sondern die eines fiktiven Deutschen mit Einflüssen des Polnischen, in dem wie bereits erwähnt zwei Fragewörter am Satzanfang stehen dürfen. Doch dann stellt sich die Frage, ob wir auch drei Instanzen von  $x^+$  haben dürfen, oder gar fünf oder zwölf. Dies ist zuvororderst eine empirische Frage; nach unserem momentanen Wissensstand scheint die maximale Anzahl an identischen Merkmalen pro Wort auf drei limitiert zu sein. Ähnlich dem Bohrschen Schalenmodell des Atoms in der Chemie benötigen wir in der Syntax also eine erweiterte Theorie, welche Merkmalskonfigurationen ein stabiles Wort ergeben. Interessanterweise sind die bisher vorhandenen mathematischen Resultate zu Minimalistischen Grammatiken vollkommen unabhängig von solchen Restriktionen. Egal wie man die Theorie in Zukunft verfeinern mag, frühere Einsichten werden ihre Gültigkeit behalten. Was aber sind nun einige dieser Einsichten, und was sagen sie uns über Sprache?

genheit bekommt, passiv Gespräche zwischen Erwachsenen zu belauschen, verkuimmert der Sprachsinn und kann später nicht mehr reaktiviert werden. Auch hierzu gibt es tragische Beispiele. Das bekannteste ist wohl Genie, die unmittelbar nach ihrer Geburt von ihren Eltern an ein Babytöpfchen gefesselt in ihr Zimmer gesperrt wurde und dort komplett von ihrer Umwelt abgeschnitten aufwuchs. Der Vater verbot seiner Familie, sich mit Genie zu unterhalten, und schlug das Mädchen, wann auch immer es Artikulationsversuche unternahm. Als sie schließlich 1970, im Alter von 13 Jahren, gefunden wurde, beherrschte Genie nur ein dutzend Wörter und ein paar Phrasen wie „Stop!“ und „Still!“. Selbst nach Jahren intensiven logopädischen Trainings konnte sie nur bruchhaft Englisch sprechen und verstehen.

Bei neurologischen Scans zeigte sich auch, dass Genie primär die rechte Gehirnhälfte zur Sprachverarbeitung verwendet, während bei einem gesunden (rechtsshändigen) Erwachsenen die linke Gehirnhälfte die dominiante ist. Die rechte Gehirnhälfte wird mitunter mit allgemeinem Problemlösen in Verbindung gebracht. Offenbar ging Genie Sprache wie ein Problem an, das sie mit non-spezifischen Strategien zu lösen versuchte. Aber gesunde Menschen verwenden für Sprache hochspezialisierte Hirnregionen, die vollautomatisch operieren. Für gesunde Menschen ist Sprache ein *Instinkt*.

Es gibt aber dennoch einen signifikanten Unterschied zwischen Instinkten und menschlicher Sprache: Wir alle sehen gleich (über den mit dem altersbedingten Zerfall einhergehenden Verlust an Sehschärfe wird hier elegant hinweggesehen), wir alle laufen gleich, aber wir sprechen nicht alle die selbe Sprache. Kinder lernen die Sprache ihrer Eltern und Freunde. Spracherwerb ist also komplizierter.

Nehmen wir zum Beispiel — um bei unserem obigen Vergleichsfall, dem Sehsinn zu bleiben — das Problem automatischer Gesichtserkennungssoftware. In den Siebziger-Jahren dachte man, dies sei ein triviales Problem, schließlich ist es ein leichtes, einem Computer hochauflösende Bilder seiner Umgebung zu liefern, und selbst die einfachsten Säugtiere exzellieren darin, ihr Umfeld zu strukturieren und Objekte anhand gewisser Merkmale zu unterscheiden. Wie sich jedoch herausstellte, ist Gesichtserkennung ein enorm schwieriges Problem, das erst 30 Jahre später soweit gemeistert wurde, dass die zuständigen Algorithmen in Endnutzeranwendungen wie Google Picasa Einzug finden konnten.

Im Vergleich zu Sprache aber ist Gesichtserkennung eine Fingerübung: Wenn wir einem Computer Gesichtserkennung beibringen wollen, können wir ein Programm für genau dieses Problem schreiben. Soll er sich allerdings mit seinem Benutzer in dessen Muttersprache unterhalten, egal um

welche Sprache es sich handelt, können wir angesichts der Unzahl gesprochener Sprachen (die sich noch dazu laufend verändern) nicht einfach ein Programm schreiben, das alle Sprachen spricht. Vielmehr benötigen wir eine Art Meta-Programm, das für jede Sprache vollautomatisch das richtige Programm schreibt. Das ist ein gewaltiges Projekt, an das sich kein Programmierer je heranwagen würde — Computerlinguisten brechen Sprache auf das Problem der Gesichtserkennung herunter, indem sie ein Programm für eine spezifische Sprache schreiben. Sie tun das, weil das Meta-Programm unbeherrschbar scheint. Aber Millionen von Kinder führen genau dieses Meta-Programm tagtäglich aus.



Zusammenfassend können wir also sagen, Sprachfähigkeit bezeichnet die natürliche Anlage des Menschen, die Sprache seiner Umgebung ohne explizite Instruktion zu erwerben und zu seinem Zwecke zu nutzen. Aber der Linguist will es natürlich genauer wissen und fragt danach, wie Sprache denn nun im Gehirn verarbeitet wird, welche Berechnungen wir unbewusst tagtäglich während unserer Pausenpläuschchen abarbeiten, was für eine Art von Wissen Sprache ist und wie man es formal-symbolisch darstellen kann, ob es strukturelle Eigenschaften gibt, die allen Sprachen gemein sind, und noch vieles mehr. Über diesen Umweg als Datenlieferant für unsere Fragen über die Sprachfähigkeit rücken dann auch die individuellen Sprachen wie Englisch und Deutsch wieder ins Rampenlicht — Sprachen sind insofern von Interesse, als sie uns ermöglichen, mehr über den Status von Sprache in unserem Gehirn zu erfahren.

## 2 Und der mathematische Linguist?

Der mathematische Linguist versucht, die im vorherigen Abschnitt aufgeworfenen Fragen unter Verwendung mathematisch ausgearbeiteter Modelle zu beantworten. Dabei gibt es zwei Ansätze: Entweder man fängt bei Null an und macht sich direkt an die mathematische Untersuchung bestimmter Probleme, oder man nimmt eine bestehende linguistische Theorie als Ausgangspunkt, untersucht diese auf mathematischer Ebene und sieht dann, welche mathematische Eigenschaften auf Sprache zutreffen unter der Annahme, dass die linguistische Theorie korrekt ist. Im ersteren Fall wird Mathematik also direkt mit Sprache in Verbindung gesetzt, während in zweiterem eine Theorie als Bindeglied zwischengeschaltet ist.



sind aber nicht nur fauler, ihre negativen Vertreter leiden obendrein an einem pathologischen Geltungsdrang — verschmilzt ein negatives Lizenzierungsmerkmal mit seinem Gegenstück, sprengt es alle Stellungsbrücken, die zuvor von anderen negativ geladenen Merkmalen auf dem selben Wort errichtet worden sind (die verschmolzenen Merkmale werden dabei nicht wieder hergestellt und haben ihr Dasein somit umsonst verwirkt). Für die Wortreihenfolge hat das zur Folge, dass das Wort, dessen Stellungsbrücken zerstört wurden, näher an die Position des Zielworts der neu errichteten Stellungsbrücke rutscht, wobei es sämtliche Wörter mitzieht, die noch immer mit ihm durch Stellungsbrücken verbunden sind.

Ein weitere Besonderheit von Merkmalen ist, dass sie innerhalb eines Wortes einer nach dem anderen wie in einer Warteschlange stehen müssen — ein Wort gleicht also einer gespannten Schnur, auf der Merkmale wie Perlen aneinandergereiht sind. Wir können also vom ersten, zweiten, dritten, ... Merkmal eines Wortes sprechen. Darüber hinaus muss jedes Wort ein Kategoriemerkmal negativer Polarität aufweisen. Eine Verbindung von Wörtern — die wir im Sinne unserer Chemie-Metapher als *Wortmolekül* bezeichnen können — ist grammatisch genau dann wenn alle Merkmale außer einem einzelnen negativen Kategoriemerkmal mit ihrem Gegenstück verschmolzen sind.

Das oben angesprochene Verhalten deutscher Fragewörter lässt sich zum Beispiel wie folgt erklären. Wir nehmen zuerst an, dass deutsch drei Kategoriemerkmale hat,  $n$  für Nomen wie *Hans* und die Fragewörter *was* und *wem*,  $v$  für Verben (= Tun-Wörter), im konkreten Fall also *gegeben*, und  $h$  für Hilfsverben wie *hat*. Das Verb *gegeben* hat drei  $n^+$ -Merkmale, die es ihm ermöglichen, Bedeutungs- und Stellungsbrücken mit *Hans*, *was* und *wem* auszubilden. Parallel hierzu trägt *hat* ein  $v^+$ -Merkmal, sodass die entsprechenden Brücken zum Verb ausgebaut werden. Es trägt aber auch ein positives Lizenzierungsmerkmal  $x^+$ . Findet sich das negative Gegenstück  $x^-$  auf einem der Nomen, baut dieses eine Stellungsbrücke zu *hat* und sprengt gleichzeitig die Stellungsbrücke zu *gegeben*. Dadurch rückt es dichter an *hat* heran und nimmt letztendlich die erste Position im Satz ein. Da *hat* aber nur ein einziges  $x^+$  hat, kann auch nur eines der Nomen eine Stellungsbrücke zu ihm ausbauen, und deshalb sind (8d) und (8e) unmöglich. Wer es ganz genau wissen will, der findet hier die Liste aller relevanten Wörter:

<i>Hans</i> : $n^-$	<i>Hans</i> : $n^- x^-$
<i>was</i> : $n^-$	<i>was</i> : $n^- x^-$
<i>wem</i> : $n^-$	<i>wem</i> : $n^- x^-$
<i>gegeben</i> : $n^+ n^+ n^+ v^-$	<i>hat</i> : $v^+ x^+ h^-$



e. \* Wem was hat Hans gegeben?

Die Grundidee der Minimalistischen Syntax zur Lösung dieses Kontrasts lässt sich mit einer chemischen Metapher umschreiben: Wörter sind Elemente, und Sätze die Moleküle, zu denen sie sich verbinden. Ein grammatischer Satz ist also sozusagen ein stabiles Molekül, ein ungrammatischer Satz ein instabiles. Aber was entscheidet darüber, ob ein Satz nun "stabil" ist oder nicht? Die Annahme ist, dass Wörter genau wie Elemente aus kleineren Teilen bestehen, und diese bestimmen, wie sich Wörter mit einander verbinden können. Im Falle der Chemie sind diese kleineren Teile Elektronen, Protonen und Neutronen, während in der Syntax schlicht von Merkmalen die Rede ist.

In Ed Stablers Formalisierung von Minimalistischer Syntax, den *Minimalistischen Grammatiken*, gibt es zwei Arten von Merkmalen, *Kategorie-merkmale* und *Lizenzierungsmerkmale*, die obendrein noch eine von zwei Polaritäten aufweisen, entweder positiv (+) oder negativ (−). Merkmale verbinden Wörter, indem sie miteinander verschmelzen. Dabei gilt es jedoch einige Regeln zu befolgen: Ein Merkmal kann nur mit genau einem Merkmal verschmelzen, und letzteres muss das gegenseitlich gepolte Gegenstück des ersten sein. Ein hypothetisches Merkmal  $x^+$  ist also kompatibel mit  $x^-$ , nicht aber mit  $x^+$  oder  $y^-$ . In gewisser Weise haben wir es also mit einem Steckausatz zu tun, bei dem ein eckiger Dübel nun mal nicht in eine rundes Loch passt.

Verschmelzen zwei Merkmale, werden die Wörter durch *Brücken* verbunden. Brücken sind eine Metapher für Abhängigkeiten zwischen Wörtern, sie haben also eine Richtung, gleich einer Einbahnstraße: das Wort mit dem negativ geladenen Merkmal ist vom Wort mit dem positiv geladenen Merkmal abhängig. Wir unterscheiden zwei Arten von Brücken: *Bedeutungsbrücken* etablieren einen rudimentären Bedeutungszusammenhang zwischen den Wörtern. Im Satz *Hans schläft*, beispielsweise, besteht eine Bedeutungsbrücke zwischen *Hans* und *schläft*, die ausdrückt, dass Hans die schlafende Person ist. *Stellungsbrücken* hingegen legen die Reihenfolge der Wörter im Satz fest. Dass *Hans* vor *schläft* steht, ist also auf eine Stellungsbrücke zurückzuführen. Wie diese Brücken nun genau in Bedeutungsrelationen und Wortreihenfolgen umgewandelt werden, würde hier zu weit führen, für unsere Zwecke reicht das Bild des Einbahnbrückenbaus durch Verschmelzung entgegengesetzter Features.

Der Unterschied zwischen Kategorie- und Lizenzierungsmerkmalen liegt nun darin, welche Brücken sie ausbilden. Kategorie-merkmale sind heißiger als Lizenzierungsmerkmale, während ihrer Verschmelzung errichten sie sowohl eine Bedeutungsbrücke als auch eine Stellungsbrücke. Im Gegensatz dazu konstruieren Lizenzierungsmerkmale nur eine Stellungsbrücke. Sie

Der erfolgreichste Vertreter der ersten Variante ist die *Theorie generativer Quantoren*, die unter anderem von Ed Keenan an der UCLA stark vorangetrieben wurde und wird. Ein Quantor ist eine spezielle Art von Artikel, der dem Nomen (= Namenwort), mit dem er sich verbindet, eine gewisse Menge, Zahl oder Proportion zuweist. Es handelt sich also um Ausdrücke wie *alle*, *keine*, *manche*. Der Begriff generalisierter Quantor erweitert diese Definition ein wenig, sodass auch Phrasen wie *alle Kinder*, *zwei Drittel von ihnen* oder *niemand* eingeschlossen werden.

Eine mathematisch natürliche Kategorie, nach der sich generalisierte Quantoren weiter unterscheiden lassen, ist jene der *Monotonizität*. Die Definition ist, wie nicht weiter verwundern dürfte, eher abstrakt. Zuerst benötigen wir ein wenig Notation: Sätze mit einem generalisiertem Quantor  $Q$  als Subjekt werden in der Form  $Qx$  geschrieben, wo  $x$  ein beliebiges Symbol ist, das den Rest des Satzes repräsentiert.

Bei einem Satz wie *Jeder Mann läuft schnell* würde  $Q$  für *jeder Mann* stehen, und für den Rest *läuft schnell* könnten wir das Symbol  $S$  wählen, sodass der gesamte Satz die Form  $QS$  hätte. Mathematisch können wir  $x$  als Menge ansehen — im konkreten Fall als Menge aller Dinge und Lebewesen, die schnell laufen. Betrachten wir stattdessen den Satz *Jeder Mann läuft*, dann wäre  $x$  die Menge aller Dinge und Lebewesen, die laufen. Nennen wir diese Menge  $L$ . Da alle Dinge, die schnell laufen, Dinge sind, die laufen, ist  $S$  eine Untermenge von  $L$ , das heißt, auf jedes Objekt, auf das  $S$  zutrifft, trifft auch  $L$  zu.

Nun können wir Monotonizität wie folgt definieren: Ein generalisierter Quantor ist *abwärts monoton* genau dann wenn im Fall, dass der Satz  $QR$  eine wahre Aussage ist, jeder Satz  $QS$ , mit  $S$  eine Untermenge von  $R$ , eine wahre Aussage ist. Er ist *aufwärts monoton*, wenn stattdessen  $R$  eine Untermenge von  $S$  ist.

Wie schon zuvor gewarnt ist die Definition ungewohnt abstrakt, sodass dem Leser jetzt vielleicht der seit Gymnasialzeiten überwunden geglaubte mathematische Angstschweiß aus den Poren schießt. Aber hinter dem ungewohnt formalen Stil versteckt sich eine simple Idee, wie die folgenden Beispiele demonstrieren. Zuvor haben wir gesehen, dass die Menge aller Dinge, die schnell laufen, eine Untermenge der Menge aller Dinge ist, die laufen. Schauen wir nun, wie sich unterschiedliche Quantoren in Bezug auf diese beiden Teilsätze verhalten.

- (3) a. Kein Mensch läuft.
- b. Kein Mensch läuft schnell.

Wenn (3a) wahr ist, dann muss (3b) auch wahr sein, denn es kann nicht sein, dass jemand schnell läuft, obwohl niemand läuft. Wir können also aus dem Wahrheitsgehalt der Obermenge “Dinge, die laufen” auf den Wahrheitsgehalt der Untermenge “Dinge, die schnell laufen” rückschließen. Darum sagen wir, dass *kein Mensch* abwärts monoton ist. Betrachten wir nun stattdessen *jeder Mensch*.

- (4) a. Jeder Mensch läuft.
- b. Jeder Mensch läuft schnell.

Hier geht der Rückschluss in die andere Richtung: Wenn jeder Mensch schnell läuft, dann läuft jeder Mensch. Somit ist *jeder Mensch* aufwärts monoton.

Aber welchen Nutzen hat dieser mathematische Hürdenlauf? Was hat eine formale Eigenschaft wie Monotonizität mit Sprache zu tun? Nun, Monotonizität spielt unter anderem eine entscheidende Rolle in der Lizenzierung von Wörtern wie Deutsch *je*. Der folgende Satz wird von deutschen Muttersprachlern ohne Zögern akzeptiert:

- (5) Kein Mensch hat je ein fliegendes Schwein gesehen.

Doch das Gegenstück mit *jeder Mensch* wird fast ausnahmslos für ungrammatisch befunden:

- (6) \* Jeder Mensch hat je ein fliegendes Schwein gesehen.

Dieses Experiment lässt sich mit unterschiedlichsten Quantoren wiederholen und führt zu folgender Generalisierung: *je* kann in Sätzen mit abwärts monotonen Subjekten erscheinen, aber nicht in Sätzen mit aufwärts monotonen Subjekten. Doch diese Regel ist nicht auf Deutsch beschränkt, sie gilt in jeder uns bekannten Sprache. Das heißt, in jeder von Menschen gesprochenen Sprache existieren Wörter wie *je*, die mit abwärts monotonen Quantoren als Subjekt auftauchen dürfen, doch nicht mit aufwärts monotonen.

Dass es zumindest eine Regel gibt, die für jede menschliche Sprache gilt, ist höchst verblüffend. Wenn Sprache im Gegensatz zum Sehsinn nicht nur von der allgemeinen Verfügbarkeit von Trainingsmaterial abhängt, sondern auch dessen spezifische Form, wieso haben sich dann Sprachen im Verlauf von tausenden von Jahren nicht soweit auseinander entwickelt, dass sie keine strukturellen Gemeinsamkeiten mehr aufweisen. Kann es sein, dass der menschliche Verstand den Formen, die Sprache annehmen kann, bestimmte Grenzen auferlegt? Doch wenn dem so ist, aus welchen Annahmen zum menschlichen Geist lässt sich die obskure mathematische Bedingung für *je* herleiten?

All das sind weitreichende Fragen, und bis heute hat niemand eine befriedigende Antwort gefunden, die über bloßes Philosophieren hinausgeht. Dabei ist das Problem sogar noch komplizierter, denn es gibt auch das Gegenstück zu *je*: Wörter wie *durchaus*, die zumindest in bestimmten Dialekten nicht mit abwärts monotonem Subjekt auftreten dürfen.

- (7) a. Jeder Mensch ist durchaus verrückt.
- b. \* Kein Mensch ist durchaus verrückt.

Solche Wörter sind seltener und die Urteile von Muttersprachlern zumindest im Deutschen weniger robust, womit sich die Frage stellt, wie sich dieser Unterschied zwischen der *je*-Klasse und der *durchaus*-Klasse erklären lässt, wo doch die beiden Regeln nahezu identisch sind und keine komplexer oder weniger natürlich scheint als die andere. Auch hier ist eine Antwort noch ausständig, doch es ist klar, dass sie nicht ohne die Mathematik auskommen kann, die zu den Generalisierungen geführt hat.



Wenden wir uns nun der zweiten Variante zu, in der Sprache nicht unmittelbar studiert, sondern bereits vorhandene linguistischen Theorien als des Sehers Rohr eingesetzt werden. Dieser Zugang ist eindeutig der populärere, sodass viele unterschiedliche Modelle existieren, mitunter auch von der gleichen Theorie. Ich werde mich hier auf *Minimalistische Grammatiken* beschränken. Sie wurden von Ed Stabler (ebenfalls UCLA) als Formalisierung *Minimalistischer Syntax* entwickelt.

*Syntax* ist der linguistische Fachbegriff für jenes Gebiet, das alle Fragen der Wohlgeformtheit von Sätzen losgelöst von Bedeutung betrifft (es ist eine kuriose Eigenschaft des Linguisten-Jargons, dass ihm stets das Kunststück gelingt, für seine Termini gängige englische Wörter zu wählen, deren direkte Pendanten in jeder anderen Sprache relativ unverständlich sind). Dass im Deutschen beispielsweise ein Fragewort wie *wer* oder *wen* am Anfang des Satzes stehen darf, aber nicht mehr als eines, während im Polnischen alle Fragewörter an den Satzanfang müssen, ist eine klassische Frage der Syntax. Mit anderen Worten, eine Theorie der Syntax muss erklären, weshalb die ersten drei Sätze im Deutschen wohlgeformt sind, der vierte und fünfte jedoch nicht.

- (8) a. Hans hat was wem gegeben?
- b. Was hat Hans wem gegeben?
- c. Wem hat Hans was gegeben?
- d. \* Was wem hat Hans gegeben?