

7. BEGLEITBLATT

Ab hier ist die Grundlage des Syntax-Skripts das Buch „Syntactic Theory“ von Ivan Sag, Thomas Wasow und Emily Bender (SWB), in der zweiten Auflage von 2003. Die Nummerierung der Skriptabschnitte versuche ich einigermaßen parallel zum Buch zu halten.

Zudem wird versucht, das HPSG-System von SWB auf das Deutsche anzuwenden, möglichst anhand von Beispielen aus dem Deutschen zu illustrieren und zugleich Unterschiede zum Englischen und anderen Sprachen herauszustellen.

5 Formale Syntax: Sag, Wasow & Bender 2003

5.1 Wozu Syntax?

- „Weil’s so faszinierend ist“. Und wer weiß: vielleicht ist’s ja auch irgendwann nützlich.
- „A Window on the Structure of the Mind“
(Innateness-Hypothese, Universal-Grammatik...)
- „A Window on the Activity of the Mind“

Ein Wunder: Obwohl Missverständnisse bei sprachlicher Kommunikation vergleichsweise selten in der Grammatik ihre Ursache haben, so zeigt sich bei isolierter Betrachtung sprachlicher Ausdrücke, dass Wörter und Sätze in hohem Grade mehrdeutig sind, ihr Bezug oft nicht eindeutig ist oder allein der sprachliche bzw. situative Kontext bestimmte Interpretationen nahelegt.

(63) lexikalische Mehrdeutigkeiten:

(B272) Hein **geht** von Poppenbüttel nach Blankenese.

(B273) Erna **geht** zum Psychiater.

(B274) Weil die Maschine nicht mehr **geht**, müssen wir rudern.

(B275) Der Verlust **geht** in die Millionen.

(B276) Die Regatta **geht** von Hamburg nach Helgoland.

(B277) Das Unternehmen **geht** den Bach runter.

(B278) Der Versuch, Atlantis zu finden, **geht** mit Sicherheit ins Auge.

(B279) Hein **geht** an die Decke.

(B280) Wie **geht**’s?

(B281) Vor der Bank steht eine Bank.

(B282) Ein Tor schießt selten ein Tor.

(64) syntaktische Mehrdeutigkeiten:

(B283) *Die Beurteilung des Studenten ist in eine Pfütze gefallen.*

(B284) *...weil der Kapitän von der Santa Maria geflohen ist.*

(B285) *Erst haben die Matrosen gemeutert, dann haben die Offiziere nachgegeben, und abschließend haben sie sich betrunken.*

(B286) *Fiete fand das Entermesser im Kopfkissen.*

(65) Indexikalische Ausdrücke:

(B287) *Ich habe hier heute noch nicht darüber geredet.*

(66) Nicht-wörtlicher Sprachgebrauch:

(B288) *Hast Du 'ne Uhr?*

(B289) *Ich hau mich aufs Ohr.*

(B290) *Fiete gibt den Löffel ab, Hein beißt ins Gras und Erna betrachtet die Radieschen von unten.*

(67) Elliptischer Sprachgebrauch & Bedeutungsrekonstruktion aus dem Kontext:

(B291) *Hein hat einen Meerjungfrauenverein gegründet. Fiete ist Vorsitzender.*

(B292) *Erna hat einen Heringssalat gemacht. Hein nascht dauernd.*

6 Some Simple Theories of Grammar**6.1 Introduction****6.2 Two Simplistic Syntactic Theories****6.2.1 Lists as Grammars**

Die einfachste „Grammatik“: eine Liste der korrekten Sätze.

Als Satz einer Sprache L gilt dann jeder Satz, der sich in der Liste befindet.

Gegenevidenz: Sätze mit iterativen Strukturen in natürlichen Sprachen, die kein definiertes Ende haben:

(B293) *Der Kapitän schläft und schläft und schläft und schläft.....zzz*

Somit:

- Listen taugen prinzipiell nicht als Grammatik natürlicher Sprachen, denn jene müssten ebenfalls unendlich sein

Ferner:

- Listen sind nicht geeignet, strukturelle Ähnlichkeiten zwischen ansonsten unterschiedlichen Sätzen angemessen zu erfassen.

- Listen sind nicht geeignet, allgemein geltende Strukturbedingungen von natürlichsprachlichen Sätzen zu erfassen.

6.2.2 Regular Expressions / reguläre Ausdrücke

Auf der Suche nach Verallgemeinerungsmöglichkeiten („Generalisierungen“) bietet es sich zunächst an, nicht die Wörter selbst, sondern ihren Typ (beispielsweise ihre Wortart) als Gegenstand syntaktischer Regeln heranzuziehen. Eine solche Grammatik bestünde dann nicht aus einer Liste von Sätzen, sondern einerseits aus einem Lexikon, das für jedes Wort dessen Typ angibt, und einer Liste von Ketten von Wortartkategorien, oder, anders gesagt, aus einer Liste von Satzmustern. Die Wortart eines Wortes ergäbe sich dabei u.a. aus seinen Distributionseigenschaften. So ließen sich immerhin Sätze wie die folgenden allesamt demselben Muster zuordnen:

(B294) *Der Steuermann trinkt.*

(B295) *Der Kapitän schläft.*

(B296) *Der Maat schnarcht.*

(B297) *Die Meerjungfrau singt.*

(68) Lexikon:

| | |
|--|---|
| <i>der, die...:</i> | D |
| <i>Steuermann, Kapitän, Maat, Meerjungfrau...:</i> | N |
| <i>trinkt, schläft, schnarcht, singt...:</i> | V |
| <i>und, oder, ...:</i> | C |

(69) Satzmuster:

D N V
D N V C V
...

Eine solche Grammatik enthielte immer noch unendlich viele Muster:

(70)

Der Kapitän schläft.

D N V

Der Kapitän schläft und schläft.

D N V C V

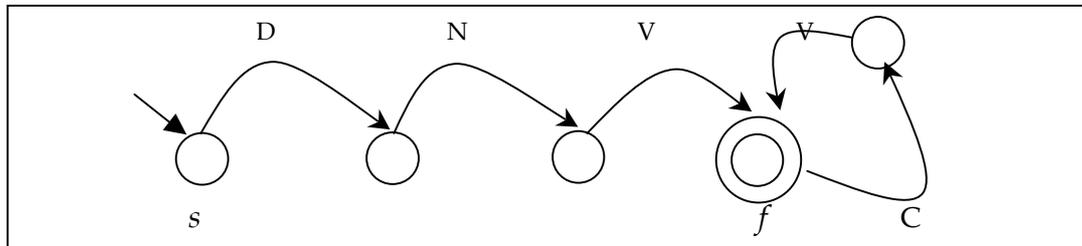
Der Kapitän schläft und schläft und schläft.

D N V C V C V. usw. usf. ad infinitum

Wenn die Möglichkeit unendlicher Iteration gegen die Wahl einer Satzlistengrammatik oder Satzmusterlistengrammatik spricht, stellt sich umgehend die Frage, welche Art Grammatik denn mit Strukturen dieser Art zurecht kommen kann. Auf den ersten Blick kommt eine sog. **Grammatik mit endlichen Zuständen** (finite state grammar) in Frage. Eine solche Gramma-

tik, als **endlicher Automat** aufgefasst, besteht aus einem „Alphabet“ (hier: die Menge der Kategoriensymbole $K = \{D, N, V, C\}$), einem Startzustand s , dem Endzustand f , einer Menge von Zwischenzuständen Z und Anweisungen, von welchem Zustand in welchen Nachfolgezustand übergegangen werden darf.

(71)



Eine derartige Grammatik gestattet es, diejenigen Muster, die sich allein durch die Wiederholung von Teilfolgen von Kategorien unterscheiden, zu einem Muster zusammenzufassen. Die null- oder n-malige Wiederholung einer Abfolge von Kategoriensymbolen wird durch den sog. **Kleene-Star** „ * “, die ein- oder n-malige Wiederholung einer Folge von Kategoriensymbolen dagegen durch das sog. **Kleene-Plus** „+“ zum Ausdruck gebracht¹⁹.

Sprachen, deren Sätze aus einer beliebig häufigen Verkettung von Bestandteilen ihres „Alphabets“ bestehen, und zwar ohne dass es Beschränkungen in Bezug auf die Anordnung der Bestandteile und die Anzahl der Wiederholungen der Verkettungen gibt, heißen **reguläre Sprachen**. Ausdrücke, die nur solche „Sätze“ (also Zeichenketten aus dem gegebenen Alphabet) beschreiben, die diesen Anforderungen genügen, heißen entsprechend **reguläre Ausdrücke**. Die in (70) bzw. 0 angedeuteten Muster lassen sich dann als regulärer Ausdruck wie folgt darstellen:

(72) $D N V (C V)^*$

Nun hat Noam Chomsky jedoch bereits 1957 gezeigt, dass auch eine Grammatik mit endlichen Zuständen keine adäquate Grammatik für das Englische ist²⁰. Ein Beispiel für einen Satz, der sich **nicht** durch eine Grammatik mit endlichen Zuständen beschreiben lässt, ist der folgende:

(B298) *Der Kapitän, der den Maat, der den Steuermann,
der die Seejungfrau(,.....) liebt, schlägt, rügt, trinkt.*

¹⁹ Auf die präzise Angabe der Bedeutung des Kleene-Stars muss hier verzichtet werden. Dies gilt ebenfalls für die gesamte Thematik der Eigenschaften formaler Sprachen, insbesondere regulärer, kontext-freier und kontext-sensitiver Sprachen sowie der Eigenschaften entsprechender Automaten im Rahmen der Automatentheorie. All dies ist Gegenstand des Seminars „Formale Sprachen und Automaten“.

²⁰ Zu den Details und dem Beweis dieser These wird hier gleichfalls auf das Seminar „Formale Sprachen und Automaten“ verwiesen.

Ferner: Problem der syntaktischen Mehrdeutigkeiten

(B299) *Wir genossen die Bootsfahrt mit dem Kapitän.*

(B300) *Auf dem Deck lagen lauter betrunkene Matrosen und Meerjungfrauen.*

(B301) *Matrosen mit Meerjungfrauen, die zuviel Rum trinken, fallen über Bord.*

⇒ Es muss in einer adäquaten Grammatik möglich sein, Abfolgen von Kategorien von Wörtern in unterschiedlicher Weise zu **Phrasen** zusammenzufassen.

6.3 Context-free Phrase Structure Grammar

Zu den Grammatiken, die Abfolgen von Kategorien zu übergeordneten Einheiten zusammenfassen können, gehören die sog. (kontextfreien) Phrasenstrukturgrammatiken (PSG).

Diese bestehen im Wesentlichen aus zwei Komponenten:

1. einem **Lexikon**, das die kategorisierten Wörter der betreffenden Sprache enthält und
2. sog. **Phrasenstrukturregeln**, die angeben, wie in der betreffenden Sprache Phrasen gebildet werden.

Phrasenstrukturregeln haben folgende allgemeine Form: $A \rightarrow \varphi$

wobei „A“ die Bezeichnung für irgendeine Phrasenkategorie ist, während „ φ “ irgendeinen regulären Ausdruck bezeichnet, gebildet aus Bezeichnungen für Wörter, Wortarten oder Phrasenkategorien. Der Pfeil „ \rightarrow “ dient in statischer Lesart als Mitteilungszeichen für die Teil-Ganzes-Relation, die zwischen denjenigen Elementen besteht, die von den Bezeichnungen auf der „rechten Seite“ des Pfeils genannt werden, und demjenigen Element, das links des Pfeiles erwähnt wird. In prozeduraler Lesart ist er als „expandiert zu“ zu lesen.

Die lexikalischen Kategorien bestehen ihrerseits aus den Wörtern, die diese Kategorie bilden. Die Kategorie N etwa aus {*Schiff, Hein, Meerjungfrau, Labskaus.....*}, die Kategorie D etwa aus {*ein, der, die, mein.....*} usw.

6.4 Applying Context-Free Grammar

Beispielgrammatik in Kapitel 2:

(73) PSG-Regeln:

| | | |
|-----|---------------|-----------------------|
| S | \rightarrow | NP VP |
| NP | \rightarrow | (D) NOM |
| VP | \rightarrow | V (NP) (NP) |
| NOM | \rightarrow | N |
| NOM | \rightarrow | NOM PP |
| VP | \rightarrow | VP PP |
| PP | \rightarrow | P NP |
| X | \rightarrow | X ⁺ CONJ X |

Diese Regelmenge weist einige erwähnenswerte Besonderheiten auf:

1. Die Kategorie **NOM**, die eingeführt wird, um u.a. koordinierte Nomina zu einer Einheit zusammenzufassen, die dann ihrerseits z.B. mit einer Negation eine Nominalphrase

bildet: $[[[keine]_{NEG} [[Matrosen]_N \text{ und}_{CONJ} [Meerjungfrauen]_N]_{NOM}]_{NP}]_{NP}$

2. Die Regel: $X \text{ ----> } X^+ \text{ CONJ } X$

Die Variable X bezeichnet hier ein beliebiges Element aus der Menge der Phrasenkategorien: {S, NP, NOM, VP, PP}. Das „+“ ist ein Kleene-Plus.

3. Bei einigen Regeln erscheint die Kategorie im Regel-Antecedens („links vom Pfeil“) erneut im Regel-Consequens („rechts vom Pfeil“). Dies sind sog. **rekursive Regeln**. Diese erlauben im Prinzip unendliche Einbettungen.

Lexikon:

- D: {die, dem, lauter, zuviel}
- A: {betrunkene}
- N: {wir, Bootsfahrt, Kapitän, Deck, Matrosen, Meerjungfrauen, Rum, Bord}
- V: {genossen, lagen, trinken, fallen}
- P: {mit, auf, über}

Aufgabe 13: Repräsentieren Sie die syntaktischen Strukturen für die Beispielsätze (B299), (B300) und (B301) durch einen Baumgraphen. Wie müsste die Regelkomponente erweitert werden?