

## Die Grammatik einer Sprache als Grundlage für die Sprachverarbeitung

*Maria Theresia Rolland*

### Zusammenfassung

Wenn man die Struktur einer Sprache vom Aspekt der Semantik her untersucht, lassen sich die Möglichkeiten der semantisch sinnvollen Verknüpfungen zwischen den Wörtern aufweisen, so daß man das immanente Beziehungsgefüge der Sprache erhält, die sogenannte semantikorientierte Grammatik. Diese dient, da sie umfassend ist, für alle Anwendungsbereiche. Es folgen Hinweise für den Aufbau entsprechender Computersysteme.

### 1 "Die empirischen Gegebenheiten"

Jeder, der sich im Bereich der Sprachverarbeitung (language processing) auskennt, weiß, daß es eine große Anzahl von Grammatiken bzw. zahlreiche Grammatiktypen gibt (Hellwig, 1989), mit deren Hilfe es möglich ist, den Computer zumindest in kleineren Anwendungsgebieten als nützliches Hilfsmittel einzusetzen. Eine generelle Verwendbarkeit des Computers im Bereich der Sprachverarbeitung dagegen ist bisher auf Grund der nicht voll durchschauten Sprachstruktur und der dadurch bedingten unzureichenden Kenntnisse über die Grammatik nicht möglich gewesen. So betont Roland Hauser (In: Tölke/Hoepfner, 1995: 15): "Nach meiner Überzeugung sind die Grammatiken nur deshalb so kompliziert, weil sie von Voraussetzungen ausgehen, die die empirischen Gegebenheiten noch nicht so richtig beschreiben."

Dem ist voll zuzustimmen - und nicht nur das. Es kann jetzt auch bewiesen werden, daß Hauser mit seiner Überzeugung recht hat; denn sobald die Struktur einer Sprache exakt erforscht vorliegt, liegt damit zugleich die Grammatik dieser Sprache vor - und kann umfassend in allen Anwendungsbereichen als Grundlage für die Verarbeitung dieser Sprache mit dem Computer dienen, besonders gut geeignet für den Aufbau von Systemen zur Wissensabfrage bzw. maschinellen Übersetzungssystemen.

## 2 Der semantikorientierte Ansatz

Der entscheidende Gesichtspunkt, warum für die jeweilige Sprache noch keine umfassende Grammatik vorliegt, ist also genau der, daß die tatsächlichen "Gegebenheiten" dieser Sprache noch nicht adäquat erforscht sind. Von einem neuem *methodischen* Ansatzpunkt aus, der das Wort und seinen Inhalt und damit die Semantik in den Mittelpunkt der Betrachtung stellt (Weisgerber, 1962; Rolland, 1994; Rolland, 1995a), ist es möglich geworden, die Prinzipien für den Aufbau solch einer Grammatik, wie Hauser sie fordert, für die deutsche Sprache zu ermitteln; d.h. in dem Moment, in dem man die wirklichen "Gegebenheiten" einer Sprache aufgedeckt hat, benötigt man nicht mehrere Grammatiken für jeweils bestimmte Anwendungsbereiche, sondern es genügt die Feststellung der *einen* Grammatik, die, da sie umfassend ist, jeder Äußerung in dieser Sprache zugrunde liegt. Damit ist die von Hauser anvisierte "solid solution" (In: Tölke/Hoepfner, 1995: 14f.) hundertprozentig möglich.

Solche grammatischen Neuheiten für das Deutsche sind z. B., um einige herauszugreifen (vgl. detailliert: Rolland, 1994), folgende:

. Es gibt nicht, wie es im Duden heißt (1995, §121ff.), 9 Wortarten, sondern 6 in sich untergliederte Wortarten.

- Es gibt nicht 3 Hilfsverben im Deutschen, sondern 4.
- Es gibt nicht 5 Satzglieder, sondern 6.
- Es gibt nicht nur die 3 Personen Singular und Plural, sondern 4

*Subjektkategorien* im Singular und Plural.

- Man muß unterscheiden zwischen dem Adjektiv (z.B. Das *schöne* Buch) und dem Adverb (Das Buch ist *schön*). Es handelt sich also um Wörter *verschiedener Wortarten*.
- Es gibt keinen z.B. Genitivus subjektivus, possessivus usw., d.h. es handelt sich bei dem semantischen Unterschied nicht um *Funktionen eines Kasus*, sondern um *Relationen zwischen Wörtern* (Rolland, 1996a).

So wird ohne weiteres deutlich, daß die semantikorientierte Untersuchung der Sprache bedeutende Einschnitte in die bisherigen Verfahrens weisen bringt. Da aber durch die neuen Erkenntnisse die Sprache in ihrer Struktur so erfaßt ist, wie sie dem Menschen wirklich vorliegt, wenn er spricht, ist sie auch exakt abbildbar und kann dementsprechend einer adäquaten Verarbeitung durch den Computer dienen. Die grundlegende Sprachstruktur ist die: Jedes Wort einer Sprache kann auf Grund seines Inhalts andere Wörter in einer bestimmten Beziehung (Relation) von sich abhängig machen, die dann auch eine bestimmte Flexion aufweisen müssen, z. B.:

*Ausgangswort:* stattfinden:

*Relation:* was, welche Sache?

*Wörter:* die Sitzung, die Tagung, die Konferenz, usw.

*Relation:* wann, zu welchem Zeitpunkt?

*Wörter:* um 10 Uhr, um 10.15 Uhr, um 11 Uhr, usw.

*Relation:* wann, während welchen Zeitraums?

*Wörter:* morgen, gestern, heute, usw.

*Relation:* wo, in welchem Raum?

*Wörter:* im Besprechungsraum, im Hörsaal 10, im Büro, usw.

*Relation:* mit welcher Bestimmtheit?

*Wörter:* möglicherweise, mit Sicherheit, vielleicht, usw.

Es handelt sich sowohl bei den Relationen, als auch bei deren Konkretisierungen um eine endliche, feststellbare Anzahl. Sind für alle Wörter (zunächst des Kernwortschatzes) diese Relationen sowie ihre Konkretisierungen als Möglichkeiten der sinnvollen Verknüpfung aufgewiesen, dann liegt diesem potentiellen Relationsgefüge der Sprache die semantikorientierte Grammatik vor. Jede Äußerung ist ein Extrakt aus diesen Vorgaben, z.B.:

Die Konferenz findet möglicherweise im Besprechungsraum statt.

Die Sitzung fand um 11 Uhr im Büro statt.

Die Sitzung wird morgen im Hörsaal 10 stattfinden.

Es geht also in erster Linie um Relationen, nicht um Satzglieder; denn unterschiedliche Relationen können durchaus gleichartig sein als Satzglied, z.B.:

- das Auto des Ministers, *Relation:* das Auto wessen, welchen Besitzers?  
*Satzglied:* Genitivobjekt
- die Rede des Ministers, *Relation:* die Rede wessen, die wer hält?  
*Satzglied:* Genitivobjekt

### 3 Sprachverarbeitende Systeme

Mit Kennungen versehen, stellt dieses Relationsgefüge als sogenannte *Relationsbasis* die Grundlage für die *automatische Wissensakquisition* dar; denn Ausgangssatz und Analyse zusammen, die per Programm erfolgt, werden als solche eingespeichert und ergeben die *Wissensbasis*. Diese dient ihrerseits als Grundlage für die *Abfrage*. Innerhalb der Analyse stellen die Relationsangaben in Form von Fragewörtern die für die Abfrage wesentliche Komponente dar; denn auf diese Weise ergeben Aussagesatz und Fragesatz echte *Pendants*, z.B.:

Was *findet mit welcher Bestimmtheit wo* *statt?*  
 Die Konferenz *findet möglicherweise* *im Besprechungsraum statt.*

Entscheidend ist, daß das Ausgangswort, hier: *stattfinden*, immer konkret sein muß. Es gibt also nicht "Baupläne" für *mehrere* Wörter, da jedes Wort sowohl die Relationen, als auch die in diesen Relationen möglichen Wörter impliziert. Diese sind jeweils *ausgangswortspezifisch*.

Anders als bisher wird also nicht *Wissen modelliert*, sondern auf der Basis der relationalen Verknüpfungsmöglichkeiten einer Sprache, also der semantikorientierten Grammatik, werden Sätze bzw. Texte, die die Wissensträger darstellen, analysiert, wobei sich zur Eingabe, wie betont, das zugehörige *Pendant* ergibt. Damit kann man, z. B. in einem Wissensabfragesystem, mit Hilfe von Fragesätzen das abrufen, was in Form von Aussagesätzen vorliegt. Das geschieht dadurch, daß dem Benutzer Wörter (Verben, Substantive) angeboten werden, aus denen er sich den gewünschten Sachverhalt, zu dem er etwas wissen möchte, in Form von Fragewörtern (mit konkreten *Ausgangswörtern*) zusammenstellt.

Für den Aufbau von *Dialogsystemen*, aus denen der Benutzer z.B. enzyklopädisches Wissen abrufen will, sind die beschriebenen Grundlagen bestens geeignet (vgl. Rolland, 1994). Die Anwendung von weitreichendster Bedeutung dürfte die *maschinelle Übersetzung* sein (Rolland, 1995b; 1996b).

## Literatur

- Duden 51995: Duden. Die Grammatik. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 0 = Duden Band 4
- Hellwig, P. 1989: Parsing natürlicher Sprachen. Grundlagen /Realisierungen. In: Batori, I. S.; Lenders, w.; Putschke, W. (Hrsg.): Computerlinguistik. Ein internationales Handbuch zur computergestützten Sprachforschung und ihrer Anwendungen. Berlin, New York: de Gruyter, 348-377/ 378-432 = Handbücher zur Sprach- und Kommunikationswissenschaft 4
- Rolland, M. Th. 1994: Sprachverarbeitung durch Logotechnik. Sprachtheorie, Methodik, Anwendungen. Bonn: Dümmler
- Rolland, M. Th. 1995a: Sprachverarbeitung auf der Basis der semantikorientierten Grammatik. In: LDV-Forum 12 ( 1995) 1, S.9-27
- Rolland, M. Th. 1995b: Fully Automated Natural Language Processing. In: ERCIM NEWS (1995) 22, S.21
- Rolland, M. Th. 1996a: Funktionen von Formen? In: Wirkendes Wort 46 (1996) 1, S.93-105
- Rolland, M. Th. 1996b: Maschinelle Übersetzung auf der Basis der Logotechnik. In: LDV-Forum 13 (1996) 1/2, S.54-76

- Tölke, S.; Hoepfner, W. 1995: Maschinen lernen sprechen - Die Computerlinguistik. Programm Bayern 2, Dienstag 6. Juni 95, 19:30 Uhr.  
In: LDV-Forum 12 (1995) 2, S. 10-16
- Weisgerber, Leo: 1962: Grundzüge der inhaltbezogenen Grammatik.  
Düsseldorf: Schwann = Von den Kräften der deutschen Sprache I (1971)