

Die semantische Gliederung als adäquate semantische Repräsentationsebene für einen sprachverstehenden 'Grafikeditor'

Johannes Müller, Holger Stahl

Zusammenfassung: In einem sprachverstehenden System soll der Bedeutungsinhalt einer gesprochenen Äußerung, die keinen Nebensatz enthält, mittels der semantischen Gliederung dargestellt werden. Diese setzt sich baumartig aus semantischen Untereinheiten (Semunen) zusammen. Sie ermöglicht es, innerhalb eines rein stochastischen 'Top-Down'-Ansatzes die Wahrscheinlichkeiten für die Emission von bestimmten Wortketten abzuschätzen. Für die Domäne 'Grafikeditor' werden Kriterien zum Erstellen eines geeigneten Typen- und Wertinventars aufgezeigt und anhand von Beispielen diskutiert. Zur Abschätzung der Auftrittswahrscheinlichkeit einer bestimmten semantischen Gliederung muß ein semantisches Modell durch Training über viele vorliegende semantische Gliederungen generiert werden. Hierzu wurden mittels einer "Wizard of Oz"-Simulation sprachliche Äußerungen gesammelt, korrespondierende semantische Gliederungen manuell erstellt und damit das semantische Modell trainiert. Ein wichtiges Kriterium ist dabei die konsistente semantische Repräsentation des Trainingsmaterials.

Schlüsselworte: Spracherkennung, Sprachverstehen, semantische Gliederung, natürlichsprachlicher Mensch-Maschine-Dialog, stochastische Modelle

1 Systemüberblick

Mit einem Grafikeditor kann ein Benutzer einfache dreidimensionale Objekte wie Kegel, Kugel, Quader oder Zylinder erzeugen, verändern oder löschen. Der komplette Mensch-Maschine-Dialog läuft dabei ausschließlich in gesprochener, deutscher Sprache ab. Das bedeutet, der Benutzer verwendet zur Eingabe weder Tastatur noch Maus, sondern nur seine Stimme. Andererseits werden Systemmeldungen nicht auf dem Bildschirm, sondern über synthetische Sprache ausgegeben.

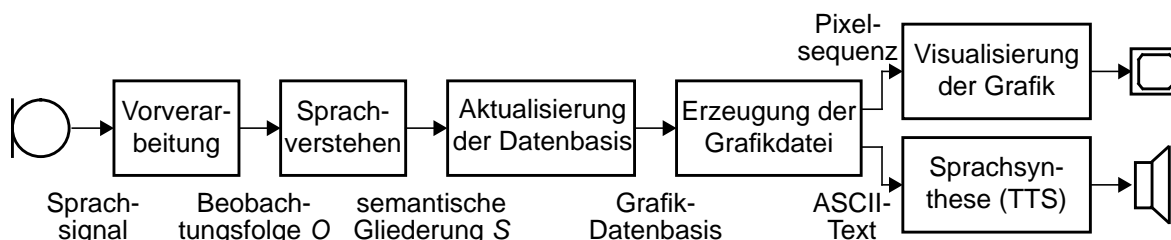


Abb. 1: Blockdiagramm des sprachverstehenden 'Grafikeditors'

Abb. 1 zeigt die Applikation im Überblick. Die folgenden Ausführungen beziehen sich auf den Block "Sprachverstehen" und auf die semantische Gliederung S.

2 'Top-Down'-Ansatz zum Verstehen natürlicher Sprache

Das Verstehen von Sprache kann aufgefaßt werden als Abbildung einer Folge von Merkmalsvektoren O [Lang 1994] auf den Bedeutungsinhalt S der zugrundeliegenden Äußerung. Für die gegebene Beobachtungsfolge O muß also aus der Menge von allen möglichen S dasjenige S_E gefunden werden, welches am wahrscheinlichsten ist. Die zu maximierende a-posteriori-Wahrscheinlichkeit $P(S|O)$ läßt sich mit dem Satz von Bayes umformen [Pieraccini 1993]:

$$P(S|O) = \frac{P(S, O)}{P(O)} = \frac{P(O|S) \cdot P(S)}{P(O)} = \sum_W \frac{P(O|W) \cdot P(W|S) \cdot P(S)}{P(O)} \quad (1)$$

Die direkte Bestimmung von $P(O|S)$ ist aufgrund der Vielfalt möglicher Kombinationen aus S und O nicht möglich. Deshalb wird die Wortebene W als weitere Repräsentationsebene eingeführt. Da für die Ermittlung von S_E nur der wahrscheinlichste Pfad interessiert [Viterbi 1973], entartet die Summe über W zum Maximumoperator. $P(O)$ muß bei der Maximierung nicht berücksichtigt werden, da es bei gegebener Beobachtungsfolge O konstant ist:

$$S_E = \operatorname{argmax}_S P(S|O) = \operatorname{argmax}_S \max_W [P(O|W) \cdot P(W|S) \cdot P(S)] \quad (2)$$

Dabei werden die Wahrscheinlichkeiten $P(O|W)$, $P(W|S)$ und $P(S)$ ausschließlich mittels stochastischer Methoden bestimmt [Müller 1994] [Stahl 1994].

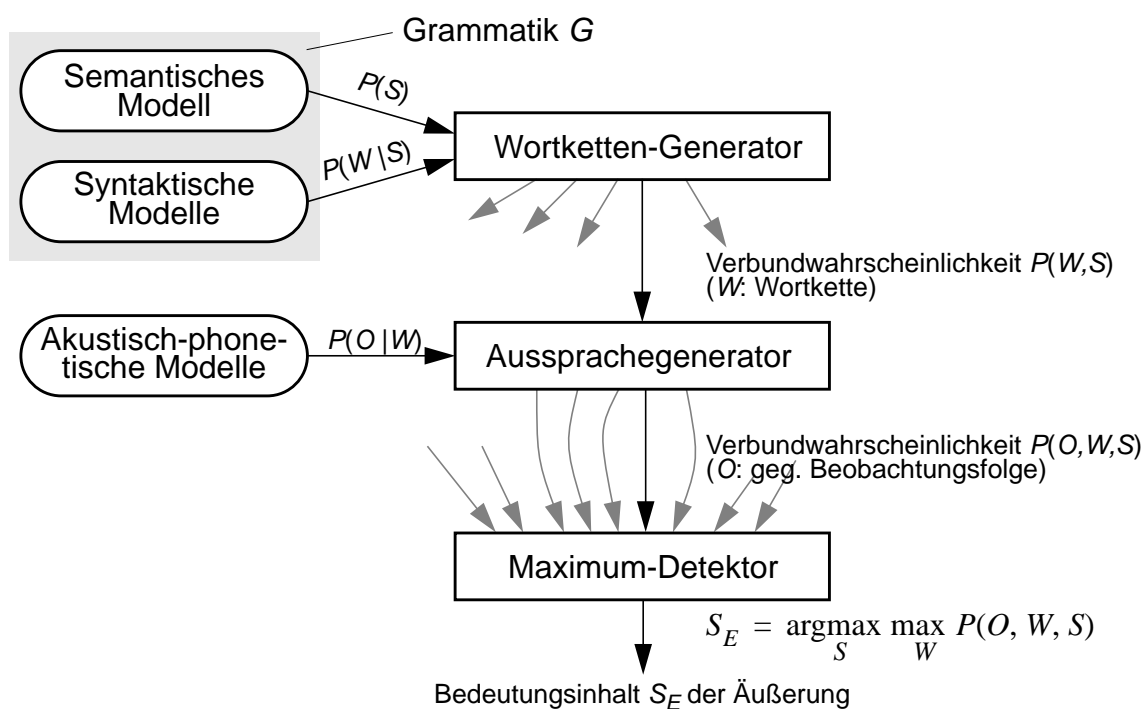


Abb. 2: 'Top-Down'-Hierarchie eines sprachverstehenden Systems

Abb. 2 zeigt die wichtigsten Module des 'Top-Down'-Erkenners, wobei innerhalb jedes dieser Module noch weitere Repräsentationsebenen (z.B. Phonemebene)

eingefügt sein können. Das beschriebene semantische Modell dient als stochastische Wissensbasis zur Bestimmung der a-priori-Wahrscheinlichkeit $P(S)$ für das Auftreten einer bestimmten semantischen Gliederung S .

3 Die semantische Gliederung

Die semantische Gliederung S kann als semantische Repräsentation einer gesprochenen Äußerung, die aus einer eng umgrenzten Domäne (in unserem Fall der 'Grafikeditor') stammt und keinen Nebensatz aufweist, aufgefaßt werden. Sie ist eine baumartige und damit hierarchische Struktur, die sich aus kleineren bedeutungstragenden Einheiten, welche im folgenden semantische Untereinheiten (oder kurz *Semune*) genannt werden, zusammensetzt.¹

$$S = \{s_1, s_2, \dots, s_N\} \quad (3)$$

Jedes Semun s_n kann als $(X+2)$ -Tupel durch seinen Typ $t[s_n]$, den Wert $v[s_n]$ und seine X Nachfolger-Semune $q_1[s_n], \dots, q_X[s_n] \in \{s_2, \dots, s_N, \text{leer}\} \setminus \{s_n\}$ beschrieben werden:

$$s_n = (t[s_n], v[s_n], q_1[s_n], \dots, q_X[s_n]) \quad , X \geq 1 \quad (4)$$

Das Semun s_1 bildet hierbei stets die Wurzel von S . Die wahlfrei durchnummerierten Semune s_2, \dots, s_N besitzen genau ein Vorgänger-Semun. Das spezielle Semun 'leer' hat den Typ $t[\text{leer}] = \text{leer}$, keinen Wert und keinen Nachfolger.

Folgende Abbildung zeigt als Beispiel den Baum einer solchen semantischen Gliederung S , bestehend aus $N=5$ Semunen.

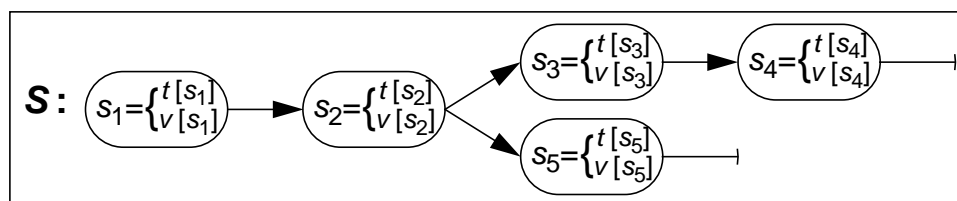


Abb. 3: Verknüpfung mehrerer Semune zum Baum einer semantischen Gliederung S

Ein bestimmtes Semun kann X direkte Nachfolger haben, verknüpft durch die Kante '—►'. Im Gegensatz dazu kennzeichnet '—|' einen leeren Nachfolger und damit das Ende eines Astes. In obigem Fall ist $X=2$ für s_2 sowie $X=1$ für alle anderen Semune (auch leere Nachfolger zählen).

Zwei semantische Gliederungen gelten als äquivalent, wenn sie die gleiche Information beinhalten. Eine äquivalente semantische Gliederung entsteht (abgesehen

¹⁾ Gegenüber früheren Darstellungen in [Müller 1994] und [Stahl 1994] ergaben sich geringfügige Änderungen der Nomenklatur.

von Beschränkungen des syntaktischen Modells) durch Vertauschen einzelner Semune innerhalb einer Sequenz von Semunen mit $X=1$ Nachfolger. In obiger semantischer Gliederung S ist somit die Anordnung der Semune s_3 und s_4 beliebig, d.h. entweder $\rightarrow \textcircled{s_3} \rightarrow \textcircled{s_4} \rightarrow$ oder $\rightarrow \textcircled{s_4} \rightarrow \textcircled{s_3} \rightarrow$. Die Anordnung der Nachfolger eines Semuns jedoch ist informationstragend und darf nicht verändert werden. So würde ein Vertauschen des Astes $\rightarrow \textcircled{s_3} \rightarrow \textcircled{s_4} \rightarrow$ mit dem Ast $\rightarrow \textcircled{s_5} \rightarrow$ die Bedeutung der semantischen Gliederung S verändern.

- Der Typ $t[s_n]$ gibt die Anzahl X der Nachfolger fest vor und schränkt die Menge möglicher Typen $t[q_1[s_n]]$, ..., $t[q_X[s_n]]$ dieser Nachfolger-Semune ein. Außerdem trifft er eine sinnvolle Auswahl möglicher ihm zuzuordnender Werte $v[s_n]$. (Derzeit werden Semune mit $1 \leq X \leq 5$ betrachtet.)
- Der Wert $v[s_n]$ gibt in der Regel die eigentliche Bedeutung des Wortes aus W an, welches dem Semun s_n zugeordnet ist.

Ein einzelnes Semun kann im weitesten Sinne als X -stellige prädikatenlogische Relationskonstante angesehen werden [Görz 1993]. Die Verknüpfung einzelner Semune zur semantischen Gliederung unterscheidet sich jedoch wesentlich von der Darstellung des Bedeutungsinhaltes mittels Prädikatenlogik. Wenn auch mathematisch nicht so exakt, bietet die semantische Gliederung folgende Vorteile:

- Die semantische Gliederung ist eine wortnahe Darstellung des Bedeutungsinhaltes. Jedem Semun s_n aus S wird genau ein bedeutungstragendes und maximal ein bedeutungsloses Wort aus der Wortkette W zugeordnet. Der Entwurf des syntaktischen Modelles (Bestimmung von $P(W|S)$) ist somit ohne weitere Repräsentationsebenen möglich.
- Zur Verknüpfung semantischer Untereinheiten gibt es nur einen Mechanismus, nämlich die Kennzeichnung weiterer Semune als sogenannte Nachfolger. Beim Entwurf der Modelle zur Berechnung der Wahrscheinlichkeiten $P(S)$ und $P(W|S)$ muß daher nur diese eine Art der Verknüpfung betrachtet werden.

Ausschließlich in der Auswahl der Typen und Werte steckt im beschriebenen Ansatz pragmatisches und linguistisches Wissen!

4 Auswahl der Typen und Werte

Pragmatischer Gesichtspunkt

Die als semantische Gliederungen darzustellenden Äußerungen sollen ausschließlich aus der Domäne 'Grafikeditor' stammen und keinen Nebensatz enthalten. Es muß daher ein geeignetes Typ- und Wertinventar gebildet werden, welches nur die zu erwartenden Äußerungen abdeckt. Für Äußerungen außerhalb der betrachteten Domäne (z.B. "heute ist schönes Wetter") werden keine Typen und Werte definiert.

Linguistischer Gesichtspunkt

Um den Suchraum beim Sprachverstehen klein zu halten und um das Auftreten möglicher Fehlinterpretationen zu minimieren, sollte das semantische Modell möglichst wenige Hypothesen bilden. Dies kann dadurch effektiv eingeschränkt werden, indem die Vielfalt möglicher Nachfolger eines bestimmten Typs gering gehalten wird. Das Problem konnte bei der Interpretation der beiden Wortketten

W_1 : mache den kegel blau

W_2 : mache den kegel groesser

beobachtet werden. Anzahl, Form, Farbe und Größe wurden zwar richtig interpretiert, doch dem Befehl 'mache' wurde fälscherweise stets derjenige Wert zugeordnet, welcher gemäß dem semantischen Modell die höchste Wertwahrscheinlichkeit besaß. Daher haben sich die anfänglich gewählten Sememe wie baf bzw. skalieren nicht bewährt.

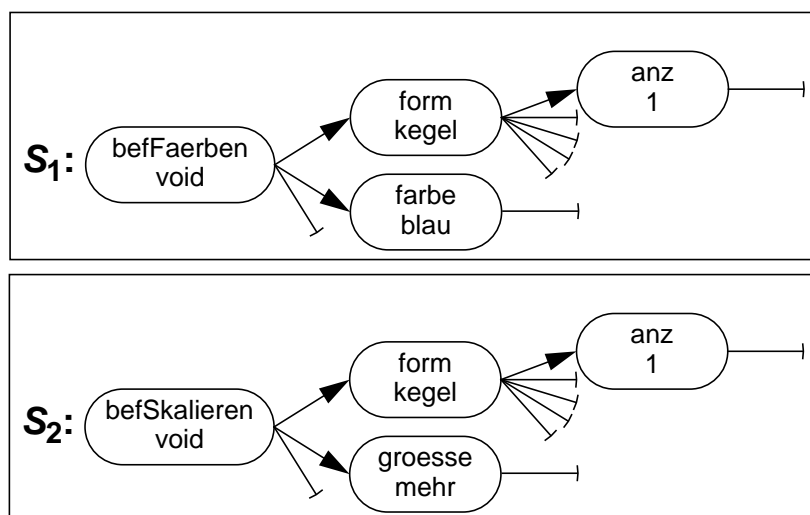


Abb. 4: Semantische Gliederungen S_1 und S_2 , welche W_1 und W_2 repräsentieren

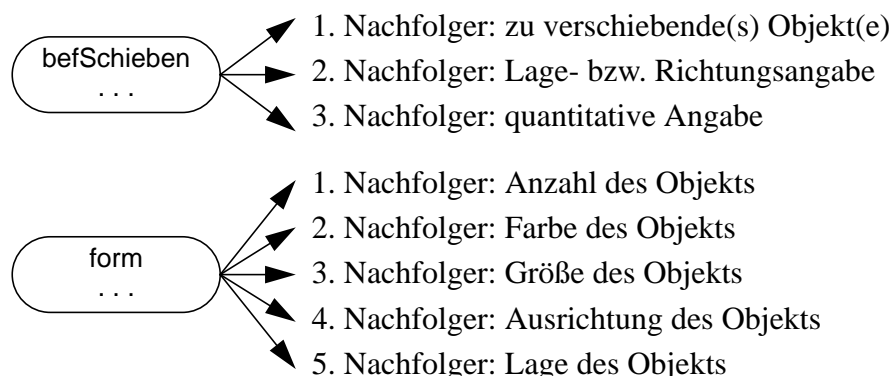
Die Folgewahrscheinlichkeit für das Auftreten einer bestimmten Nachfolgerkombination ist definitionsgemäß nur vom Typ abhängig. So ist es zweckmäßig, diese Abhängigkeit bei der Typ-Deklaration zu berücksichtigen. In obigem Beispiel sind die dem bedeutungstragenden Wort 'mache' zugrundeliegenden Sememe, welche die jeweilige Anzahl möglicher Nachfolger minimieren, baf bzw. skalieren , wobei jeweils der zweite Nachfolger mit großer Wahrscheinlichkeit eine Farbe bzw. eine Größe ist.

Derzeit werden zum Beschreiben des gesammelten Sprachmaterials aus der Domäne 'Grafikeditor' 35 Typen mit insgesamt 248 Werten eingesetzt. Tab. 1 zeigt einen Teil des derzeit benutzten Typen- und Wertinventars.

Typ	X	Werte	Erläuterung
achse	1	x, xRechts, xLinks, y, yRechts, yLinks, z, zRechts, zLinks	Achse evtl. mit Drehrichtung
anz	1	1, 2, 3, 4, ..., alle	Anzahl bestimmter Objekte
anzAllg	1	0, 1, 2, 3, ..., einige	Anzahl unbestimmter Objekte
bef	1	gruppieren, loeschen, vertauschen	Bef. für ein bestimmtes Objekt
befAllg	1	erzeugen	Bef. für ein unbestimmtes Obj.
befDrehen	3	void, waagrecht	Befehl zum Drehen
befFaerben	3	void	Befehl zum Färben
befSchieben	3	auseinander, ausrichten, beruehren, void, zentrieren	Befehl zum Verschieben
befSkalieren	3	faktor0,5, faktor2, faktor3, mehr, void, weniger	Befehl zum Skalieren
befSteuern	1	abspeichern, beschimpfen, drucken, ende, initialisieren, irrelevant, korrektur, wiederholen	Befehl zum Steuern des Programmablaufs
befUmformen	2	void	Befehl zum Umformen
farbe	1	beige, bunt, blau, braun, dunkel, dunkler, durchsichtig, gelb, grau, gruen, gold, hell, heller, oliv, rot, tuerkis, ...	Farbangabe
form	5	kegel, kugel, quader, unbestimmt, zylinder	Form eines best. Objekts
formAllg	5	kegel, kugel, quader, unbestimmt, zylinder	Form eines unbest. Objekts
groesse	1	mehr, mehrBreite, mehrHoehe, mehrTiefe, viel, vielHoehe, vielBreite, vielTiefe, wenig, wenigHoehe, wenigBreite, wenigTiefe	Größenangabe
lage	1	aussen, dort, hinten, links, mitte, oben, rechts, unten, vorne	Lagenangabe
lageBew	1	nachAussen, nachHinten, nachLinks, nachOben, nachRechts, nachUnten	Bewegung eines Objekts
lageRel	1	anstelle, hinter, innerhalb, linksNeben, nach, neben, parallel, rechtsNeben, ueber, unter, vor, zwischen	Lagenangabe relativ zu einem Objekt
logik2	2	ausser, und	logische Verknüpfung
quant	1	etwas, ganz, viel, 1mm, 2mm, 3mm, ..., 10grad, 20grad, faktor0,1, faktor0,5, faktor2, faktor3	quantitative Angabe

Tab. 1: Ausschnitt des Typen- und Wertinventars des 'Grafikeditors'

Hat ein Semun mehrere Nachfolger, ist deren Anordnung streng festgelegt. Eine Vertauschung hätte eine falsche semantische Interpretation zur Folge.



5 Mehrdeutigkeiten der semantischen Gliederung

Wie bereits vorher angesprochen, ist die semantische Gliederung eine wortnahe Darstellung. Um die an sie gestellten Randbedingungen einzuhalten (siehe Kap. 3), muß jedoch derselbe Bedeutungsinhalt unter Umständen durch unterschiedliche semantische Gliederungen ausgedrückt werden. Es seien die folgenden Wortketten betrachtet:

W_3 : mache die kugel doppelt so_gross

W_4 : vergroessere die kugel auf das_zweifache

W_5 : verdopple die_groesse der kugel

Trotz gleichen Bedeutungsinhalts weichen die korrespondierenden semantischen Gliederungen voneinander ab:

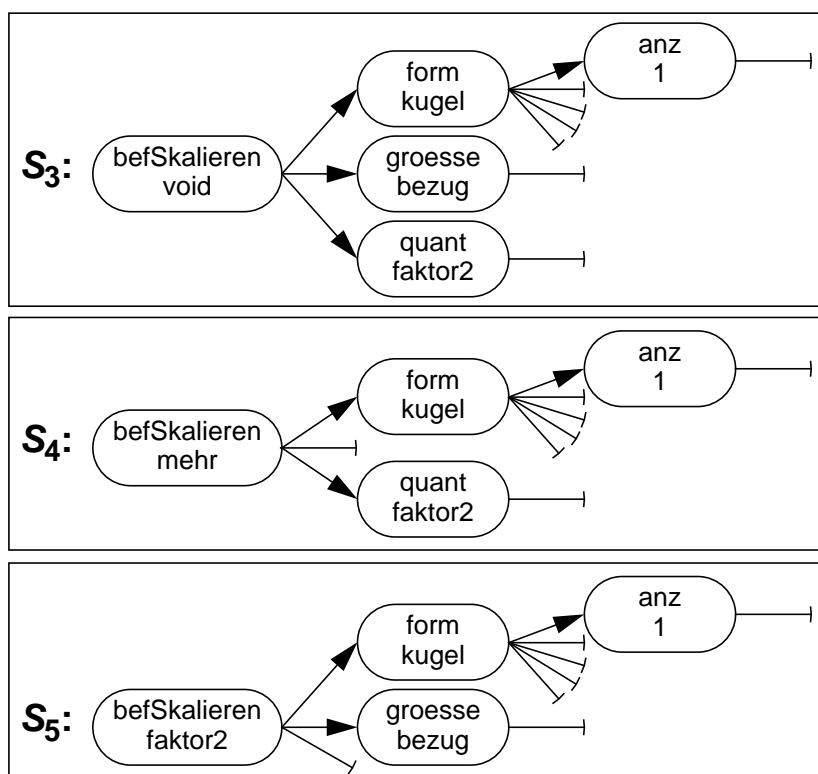


Abb. 5: Semantische Gliederungen S_3 , S_4 und S_5 korrespondierend zu W_3 , W_4 und W_5

Um dieser Mehrdeutigkeit zu begegnen, könnte die wortnahe semantische Gliederung in einem weiteren Verarbeitungsschritt regelbasiert in eine eindeutige semantische Repräsentation überführt werden.

Trotz dieser Ambiguität ist die semantische Gliederung mit der Bereitstellung von stochastischen Aussagen über zugrundeliegende Wortketten eine adäquate Repräsentationsebene in einem domänenspezifischen, sprachverstehenden System.

6 Erstellen des Trainingsmaterials

Um die entsprechenden Wahrscheinlichkeiten des semantischen und der syntaktischen Modelle zu schätzen, muß zunächst Trainingsmaterial gesammelt werden. Damit diese Abschätzung der Realität möglichst nahe kommt, sollten von vielen Sprechern jeweils viele Äußerungen herangezogen werden. Die akustisch-phonetischen Modelle zur Bestimmung von $P(O|W)$ können von bestehenden Spracherkennungssystemen (z.B. SPICOS [Höge 1990] oder SPRING [Wothke 1989]) übernommen und für hiesige Betrachtungen als gegeben angesehen werden. Da in den syntaktischen Modellen und im semantischen Modell nur die Wahrscheinlichkeiten $P(W|S)$ und $P(S)$ abgeschätzt werden, genügen also zum Training der beiden letztgenannten stochastischen Wissensbasen die Wortkette W und die semantische Gliederung S jeder Äußerung.

6.1 Sammeln sprachlicher Äußerungen mit "Wizard of Oz"-Simulation

Um möglichst authentische Äußerungen zu sammeln (ein Mensch redet anders als er schreibt), wurde mit mehreren Versuchspersonen eine "Wizard of Oz"-Simulation durchgeführt. Innerhalb dieser Versuchsanordnung simuliert ein "Wizard" (Zauberer) ohne Wissen des Benutzers die sprachverstehende Funktion des Rechners. Derartige "Wizard of Oz"-Simulationen sind u.a. in [Blomberg 1993], [Dybkjaer 1993] oder [Katunobu 1994] beschrieben. Die Aufgabe der Versuchsperson ist es, mittels Spracheingabe eine beliebige Grafik am Bildschirm zu editieren. Objekte wie Kegel, Kugel, Quader oder Zylinder können erzeugt, verändert oder gelöscht werden. Dabei wird vorgegeben, nur Anweisungen aus der Domäne 'Grafikeditor' ohne Nebensätze zu sprechen.

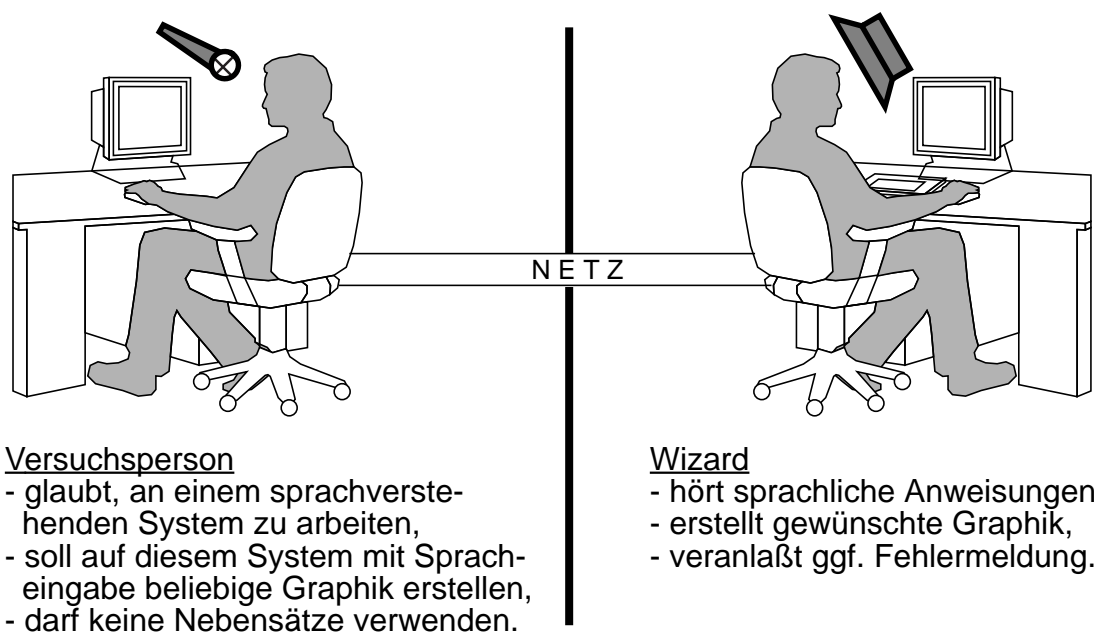


Abb. 6: Prinzip der "Wizard of Oz"-Simulation

Somit wurden von 33 Versuchspersonen insgesamt 1915 sprachliche Äußerungen gesammelt. Davon erfüllten 1843 Äußerungen die getroffenen Vorgaben.

6.2 Erstellen von Wortketten und semantischen Gliederungen

Zu jeder Äußerung wird anschließend manuell eine Textdatei erstellt. Diese enthält neben der Wortkette $W = w_1 w_2 \dots w_j \dots w_J$ die semantische Gliederung S und eine exakte Zuordnung der von jedem einzelnen Semun s_n aus S emittierten Worte w_j aus W .

Die semantischen Gliederungen der ersten ca. 700 Äußerungen mußten jeweils rein manuell erstellt werden. Mit den daraus erstellten semantisch-syntaktischen Modellen konnte dann ein Großteil der weiteren Wortketten automatisch in korrespondierende semantische Gliederungen gewandelt werden.

7 Training des semantischen Modells

Das semantische Modell muß eine endliche Anzahl Parameter enthalten, die sich einerseits zuverlässig aus begrenztem Trainingsmaterial bestimmen lassen, die andererseits jedoch auf die jeweilige a-priori-Wahrscheinlichkeit $P(S)$ einer unbegrenzten Menge semantischer Gliederungen S schließen lassen.

Da statistische Abhängigkeiten nur innerhalb eines Semuns angenommen werden, kann die a-priori-Wahrscheinlichkeit $P(S)$ als Produkt folgender bedingter Wahrscheinlichkeiten erster Ordnung berechnet werden:

$$P(S) = f_{\text{WZL}} \cdot \prod_{n=1}^N (e_n \cdot f_n), \text{ mit ...} \quad (5)$$

- ... f_{WZL} als a-priori-Wahrscheinlichkeit, daß das Wurzel-Semun s_1 vom Typ $t[s_1]$ ist:

$$f_{\text{WZL}} = P(t[s_1]) \quad (6)$$

- ... e_n als bedingte Wahrscheinlichkeit, daß ein Semun vom Typ $t[s_n]$ den Wert $v[s_n]$ aufweist:

$$e_n = P(v[s_n] | t[s_n]) \quad (7)$$

- ... f_n als bedingte Wahrscheinlichkeit, daß ein Semun vom Typ $t[s_n]$ X Nachfolger-Semune $q_1[s_n], \dots, q_X[s_n]$ vom Typ $t[q_1[s_n]], \dots, t[q_X[s_n]]$ aufweist:

$$f_n = P(t[q_1[s_n]], \dots, t[q_X[s_n]] | t[s_n]) \quad (8)$$

Initialisierung: Zuerst werden für alle semantischen Gliederungen des Trainingssets die entsprechenden Häufigkeiten H inkrementiert und am Ende durch die jeweils zutreffende Gesamthäufigkeit dividiert. So lassen sich die oben angesprochenen Wahrscheinlichkeiten abschätzen:

$$P(\tau) = \frac{H(\text{Typ} = \tau \text{ in höchster Hierarchiestufe})}{H(\text{alle } S)} \quad (9)$$

$$P(\varphi|\tau) = \frac{H(\text{Typ} = \tau, \text{Wert} = \varphi)}{H(\text{Typ} = \tau)} \quad (10)$$

$$P(\tau_1, \dots, \tau_X|\tau) = \frac{H(\text{Typ} = \tau, \text{Nachfolger-Typen} = \tau_1, \dots, \tau_X)}{H(\text{Typ} = \tau)} \quad (11)$$

Iteration: Mit dem so initialisierten Modell wird anschließend iterativ die a-priori-Wahrscheinlichkeit $P(S)$ durch Erzeugen der jeweils wahrscheinlichsten äquivalenten semantischen Gliederung (Variation der Reihenfolge innerhalb Sequenzen von Semunen mit jeweils $X=1$) maximiert. Dies wird solange wiederholt, bis sich $P(S)$ aller Äußerungen des Trainingsmaterials nicht mehr ändert.

8 Evaluierung

Die Wortketten W_{ref} mehrerer Testäußerungen werden mittels vorher trainierter semantisch-syntaktischer Modelle¹ sequentiell in Hypothesen semantischer Gliederungen S_{hyp} gewandelt. Diese Gliederungshypothesen S_{hyp} werden mit den Gliederungsreferenzen S_{ref} verglichen und daraus die Semantik- und die Semun-Erkennungsraten berechnet.

Mit identischen Trainings- und Testdaten können alle vorliegenden 1843 "gültigen" Wortketten zu 99,35% in korrekte semantische Gliederungen überführt werden. Die Semun-Erkennungsrate ist dabei 99,92%. In einer anderen Evaluierung werden die Modelle mit 1751 Äußerungen trainiert und mit 92 anderen Äußerungen getestet. Die Semantik-Erkennungsrate sinkt dabei auf 80,43%, die Semun-Erkennungsrate auf 79,39% ab. Ein wesentlicher Grund für diesen Einbruch an Erkennungssicherheit ist, daß mit dem derzeit benutzten Trainingsmaterial die resultierende Größe des Wörterbuchs (Menge aller im Trainingsmaterial vorkommender Worte) nicht ausreicht. Sobald auch nur ein Wort w_j einer Wortkette W_{ref} der Testdaten im Wörterbuch nicht existiert, versagt die 'Top-Down'-Suche des Textverstehens für die gesamte Äußerung. Daher erscheint eine Fortsetzung der "Wizard of Oz"-Versuche zur Gewinnung weiterer Äußerungen insbesondere im Hinblick auf eine Erweiterung des Wörterbuchs notwendig.

¹⁾ Auf das Training der syntaktischen Modelle wird in diesem Beitrag nicht näher eingegangen.

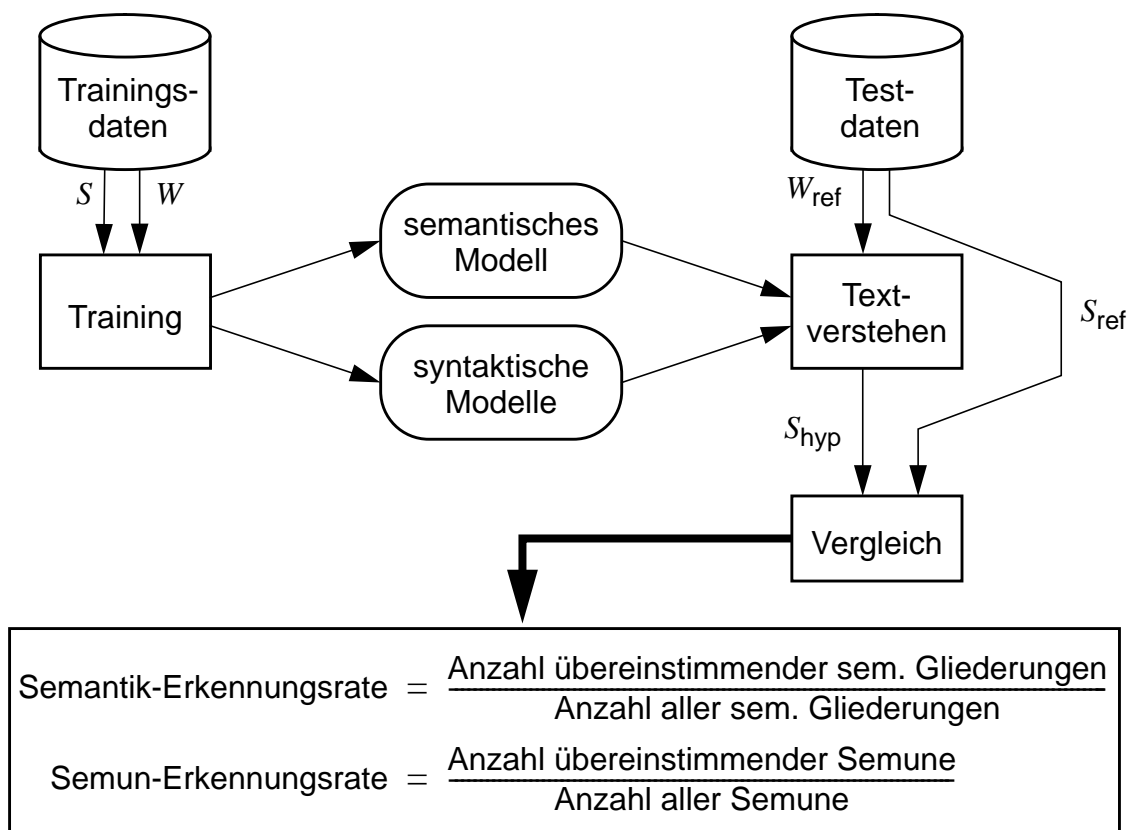


Abb. 7: Messung der Erkennungsraten

Die differenzierte Behandlung der bedingten Wahrscheinlichkeiten in den semantisch-syntaktischen Modellen haben auf die resultierenden Erkennungsraten spürbaren Einfluß. Werden in "qualitativen" Modell sämtliche vorhandenen Wahrscheinlichkeiten (d.h. $P \neq 0$) als gleichverteilt betrachtet, sinken die Semantik- und Semun-Erkennungsraten spürbar ab.

	Semantik-Erken.rate	Semun-Erken.rate
1843 identische Trainings- und Testäußerungen	99,35%	99,92%
1843 identische Trainings- und Testäußerungen "qualitatives" semantisches Modell	94,19%	99,31%
1843 identische Trainings- und Testäußerungen "qualitative" syntaktische Modelle	93,05%	99,37%
1843 identische Trainings- und Testäußerungen "qualitative" semantische und syntakt. Modelle	90,45%	99,01%
1751 Trainingsäußerungen, 92 Testäußerungen	80,43%	79,39%

Tab. 2: Erzielte Semantik- und Semun-Erkennungsraten

Die erzielten Semantik- und Semun-Erkennungsraten beweisen zum einen, daß die Wahrscheinlichkeiten in den Modellen durch Trainingsmaterial abgeschätzt werden müssen, zum anderen sind sie hoch genug, um die Erweiterung auf gesprochene Äußerungen innerhalb eines sprachverstehenden Systems zuzulassen.

9 Literatur

- Blomberg, Mats et al. (1993): *An Experimental Dialogue System: Waxholm*, Tagungsband Eurospeech 1993 (Berlin, Deutschland), S. 1867-1870
- Dybkjaer, Hans; Bernsen, Niels Ole; Dybkjaer, Laila (1993): *Wizard-of-Oz and the Trade-off between Naturalness and Recognizer Constraints*, Tagungsband Eurospeech 1993 (Berlin, Deutschland), S. 947-950
- Görz, Günther (1993): *Einführung in die künstliche Intelligenz*, Addison-Wesley
- Höge, Harald (1990): *SPICOS II - a Speech Understanding Dialogue System*, Tagungsband ICSLP 1990 (Kobe, Japan), S. 1313-1316
- Katunobu, Itou et al. (1994): *Collecting and Analyzing Nonverbal Elements for Maintenance of Dialog Using a Wizard of Oz Simulation*, Tagungsband ICSLP 1994 (Yokohama, Japan), S. 907-910
- Lang, Manfred; Stahl, Holger (1994): *Spracherkennung für einen ergonomischen Mensch-Maschine-Dialog*, Zeitschrift 'mikroelektronik', Band 8 (1994), Heft 2, S. 79-82
- Müller, Johannes; Stahl, Holger (1994): *Ein Ansatz zum Verstehen natürlicher, gesprochener Sprache durch hierarchisch strukturierte Hidden-Markov-Modelle*, Tagungsband KONVENS 1994 (Wien, Österreich), S. 260-269
- Pieraccini, Roberto; Levin, Esther; Vidal, Enrique (1993): *Learning how to Understand Language*, Tagungsband Eurospeech 1993 (Berlin, Deutschland), S. 1407-1412
- Stahl, Holger; Müller, Johannes (1994): *An Approach to Natural Speech Understanding Based on stochastic Models in a Hierarchical Structure*, Tagungsband 'Modern Modes of Man-Machine-Communication', 1994 (Maribor, Slowenien), S. 16/1-16/9
- Viterbi, A.J. (1973): *Error Bounds for Convolutional Codes and an Asymptotical Optimal Decoding Algorithm*, IEEE Trans. Information Theory, Band 61 (1973), S. 268-278
- Wothke, Klaus et al. (1989): *The SPRING Speech Recognition System for German*, Tagungsband Eurospeech 1989 (Paris, Frankreich), S. 2/9-2/12