Gefördert vom
bmb+f
Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie

Verbmobil
Verbundvorhaben

Robuste semantische Verarbeitung

Karsten L. Worm

Universität des Saarlandes

Vm

Report 200
Dezember 1997

Dezember 1997

Karsten L. Worm

Computerlinguistik
Universität des Saarlandes
D-66041 Saarbrücken

Tel.: (0681) 302 - 4503
Fax: (0681) 302 - 4351
e-mail: worm@coli.uni-sb.de

Gehört zum Antragsabschnitt: 2.5.3 Robuste semantische Verfahren

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 701 R 4 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

ISSN 0949-6084

Zusammenfassung

Dieser Report erläutert den Inhalt des Arbeitspakets *2.5.3 Robuste semantische Verfahren* im Verbmobil 2-Netzplan und die Vorstellungen der Projektpartner an der Universität des Saarlandes zu dessen Bearbeitung.

Die spontansprachlichen Äußerungen im Verbmobil-Corpus können vom Verbmobil-Forschungsprototypen 1.0 in vielen Fällen nicht vollständig überspannend syntaktisch-semantisch analysiert werden, worunter die Übersetzungsqualität und die Systemperformanz erheblich leiden. Dieses Papier beschreibt erste Ideen zur Lösung dieser Probleme.

Inhaltsverzeichnis

1 Einführung

Das Projekt Verbmobil (Bub, Wahlster, und Waibel 1997; Wahlster 1997) beschäftigt sich mit der maschinellen Übersetzung spontan gesprochener Sprache in Verhandlungsdialogen. Dabei werden im Verbmobil-System verschiedene Verfahren der maschinellen Übersetzung nebeneinander angewendet: Einerseits wird ein semantischer Transferansatz verfolgt, dem ein linguistischer Analyseprozeß bis zur Semantik der Eingabeäußerung zugrundeliegt, andererseits werden alternative Verfahren wie statistische und dialogaktbasierte Übersetzung erprobt.

In dem transferbasierten Verarbeitungsstrang wird eine gesprochene Eingabe zunächst von einem Spracherkenner auf eine Menge von gewichteten Worthypothesen abgebildet. In der Menge der Worthypothesen suchen mehrere syntaktisch-semantische Parser, die auf verschiedenen computerlinguistischen Anätzen basieren, eine analysierbare Wortkette. Das unter ihren Ausgaben ausgewählte Ergebnis, eine semantische Repräsentation der quellsprachlichen Äußerung, wird von der Transferkomponente auf eine zielsprachliche semantische Repräsentation abgebildet, aus der eine Äußerung in der Zielsprache generiert und synthetisiert wird.

Die Arbeiten in der ersten Projektphase haben gezeigt, daß eine die gesamte Äußerung überspannende syntaktisch-semantische Analyse als Eingabe für den Transfer nicht in allen Fällen zu erwarten ist, worunter die Übersetzungsqualität und die Systemperformanz leiden. Dieses Problem hat mehrere Ursachen:

Spontansprachliche Phänomene: In spontan gesprochener Sprache sind sehr oft Selbstkorrekturen und Abbrüche zu finden, die verwendeten Konstruktionen sind häufig unvollständig und ungrammatisch. Etwa jeder fünfte Turn im Verbmobil-Corpus enthält mindestens eine Selbstkorrektur oder einen Abbruch.

Erkennerfehler: Trotz erheblicher Fortschritte bei der Spracherkennung ist die Ausgabe des Spracherkenners nicht immer perfekt. Oftmals genügt ein falsch erkanntes Wort, um eine Analyse scheitern zu lassen.

Segmentierungsprobleme: Mitunter werden prosodische Grenzen an Positionen erkannt, wo die Syntax keine Grenze annimmt, oder die Syntax erzwingt eine Segmentierung der Eingabe, die den Vorhersagen der Prosodie zuwiderläuft.

Unzureichende Abdeckung der Grammatik: Die Abdeckung der Grammatik eines Systems ist in realistischen Anwendungen stets unvollständig; außerdem treten beim Wechsel zu einer neuen Domäne gewöhnlich unbekannte, aber grammatische Konstruktionen auf.

Der Verbmobil-Forschungsprototyp, das in den ersten vier Jahren (1993-96) des Projekts entwickelte System, stellt zwar bereits einen großen Fortschritt dar, was die Behandlung dieser Probleme angeht (Block 1997); sie können aber noch längst nicht als gelöst angesehen werden.

Aufgrund dieser Erfahrung aus der ersten Projektphase wurde das Arbeitspaket *2.5.3 Robuste semantische Verfahren* im Verbmobil 2-Netzplan (Burgard et al. 1996, 85f.) definiert. Das Ziel dieses Arbeitspakets ist, Verfahren zu entwickeln und zu implementieren, die partielle Analyseresultate anhand von Heuristiken zu größeren (Teil-)Analysen zusammenfügen, dabei Selbstkorrekturen und Abbrüche eliminieren und so die Qualität der semantischen Analyse und letztendlich der Übersetzung erhöhen.

Die Resultate der flachen und tiefen Analyseverfahren sollen in diesem Prozeß miteinander kombiniert werden, um die Stärken der verschiedenen Ansätze zu bündeln und ihre Schwächen auszugleichen.

Dieser Report beschreibt die Aufgaben des Arbeitspaketes und stellt erste Ideen zur Lösung der anstehenden Probleme vor. In Abschnitt 2 wird zunächst ein Überblick über die Probleme gegeben, mit denen man bei der Verarbeitung spontan gesprochener Sprache konfrontiert ist, und in Abschnitt 3 werden Ansätze aus der Literatur zur Lösung dieser Probleme referiert. Abschnitt 4 umreißt die Funktion des Moduls „Robuste semantische Verfahren“, für die in Abschnitt 5 Beispiele diskutiert werden. Abschnitt 6 faßt den Inhalt dieses Papiers zusammen.

2 Probleme der Verarbeitung von Spontansprache

2.1 Spontansprachliche Phänomene

Spontan gesprochene Sprache unterscheidet sich deutlich von geschriebener Sprache (Tropf 1994; Höhne-Leska 1975; Weiss 1975), was besondere Probleme für die machinelle Verarbeitung aufwirft (Block 1997).

2.1.1 Eigenschaften gesprochener Sprache

Spontansprachliche Äußerungen enthalten häufig

- Konstruktionen, die in geschriebener Sprache nicht vorkommen,
- fragmentarische Äußerungen,
- Selbstkorrekturen, sowie
- ungrammatische Konstruktionen.

In der Grammatik eines sprachverarbeitenden Systems lassen sich diese Phänomene nur eingeschränkt behandeln. Konstruktionen, die in geschriebener Sprache ungrammatisch, in gesprochener Sprache aber zulässig sind, sollten natürlich von der Grammatik abgedeckt werden (Block und Schachtl 1995). In diese Kategorien fallen z. B. die sehr oft vorkommenden Vorfeldellipsen.

Die Behandlung anderer spontansprachlicher Phänomene im engeren Sinne wie Abbrüche und Selbstkorrekturen in der Grammatik durch zusätzliche Regeln vergrößert aber den Suchraum extrem und schränkt so die Suche des Parsers auf den Wörtergittern zu wenig ein, so daß er u. U. einen ungrammatischen Pfad im Wörtergitter bereits akzeptiert, während auch ein grammatischer Pfad darin enthalten ist, der bei weiterer Suche gefunden würde.

Diese Phänomene müssen also auf andere Weise behandelt werden. Im Verbmobil-Forschungsprototyp wurde die tiefe linguistische Analyse hierzu durch flache Analyse- und Übersetzungsverfahren ergänzt, deren Ergebnisse im Falle des Fehlschlagens der tiefen Analyse ausgegeben werden. Diese flachen Verfahren sind zwar so robust, daß sie auf jeder Eingabe ein Resultat liefern, ihre Ergebnisse sind aber häufig inkorrekt, so daß eine Erhöhung der Robustheit der tiefen Analyse anzustreben ist.

2.1.2 Terminologie

Die Terminologie, die zur Beschreibung spontansprachlicher Phänomene verwendet wird, ist ausgesprochen uneinheitlich. Zwei Sichtweisen auf spontansprachliche Phänomene herrschen vor,

- eine *funktionale* Sicht, die die Funktion der Äußerung und ihrer Teile betont, z. B. daß ein Äußerungsteil einen vorhergehenden korrigiert, und
- eine *deskriptive* Sicht, die sich auf die strukturellen Eigenschaften einer Äußerung bezieht, z. B. daß sie ein Anakoluth enthält.

In diesem Papier werden die beiden Phänomene *Selbstkorrektur* und *Abbruch* betrachtet. Unter einer funktionalen bzw. deskriptiven Sicht genügen für die Zwecke dieses Papiers die folgenden Definitionen:

Selbstkorrektur: Der Sprecher korrigiert innerhalb eines Turns einen Teil der bisherigen Äußerung, z. B. weil er eine andere Information als die intendierte verbalisiert hat, eine andere, für der Hörer besser verständliche Formulierung wählen möchte, oder seine Intentionen während der Äußerung geändert hat (funktionale Sicht).

Eine Äußerung, die eine Selbstkorrektur enthält, kann, muß aber nicht ungrammatisch sein. An einer Stelle unterbricht der Sprecher die Äußerung und setzt sie nach einer u. U. durch Interjektionen gefüllten Pause fort und äußert etwas, was einen Teil der Äußerung vor der Unterbrechung ersetzt (deskriptive Sicht). Vgl. Abschnitt 2.1.3.

Abbruch: Der Sprecher bricht eine Äußerung ab, z. B. weil er seine Intentionen geändert hat oder die Äußerung nicht für zweckmäßig hält, und beginnt eine neue, die einen anderen Inhalt hat (funktionale Sicht).

Die abgebrochene Äußerung ist meistens ein syntaktisch unvollständiges Präfix einer Phrase; wenn sie einen vollständigen Satz bildet, muß anhand der Prosodie erkennbar sein, daß die Äußerung unvollständig ist (deskriptive Sicht).

2.1.3 Struktur von Selbstkorrekturen

In diesem Abschnitt werden einige wichtige Ergebnisse von Levelt (1983, 1989) zu Eigenschaften und Struktur der Selbstkorrekturen referiert. Auf seine Untersuchungen beziehen sich viele Arbeiten zu Selbstkorrekturen in Linguistik und Sprachverarbeitung. Zum Vergleich wird kurz das Modell von Nakatani und Hirschberg (1993) dargestellt.

Das Modell von Levelt Selbstkorrekturen sind das häufigste spontansprachliche Phänomen. Sie weisen eine gemeinsame Struktur auf; Abb. 1 zeigt die allgemeine Struktur einer Selbstkorrektur nach Levelt (1983). Nach diesem Modell besteht eine Selbstkorrektur aus drei Teilen:

1. Einer Originaläußerung (*original utterance*),

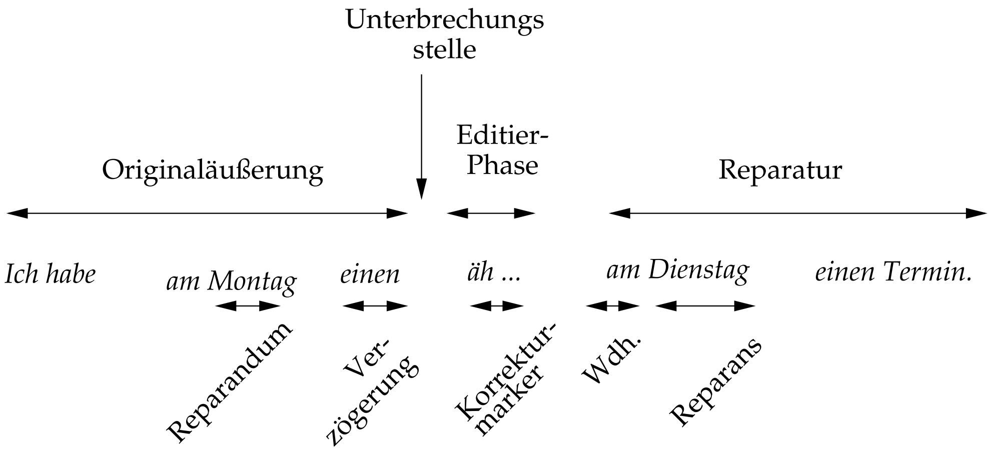


Abbildung 1: Struktur einer Korrektur nach Levelt (1983)

2. einer Editierphase (*editing phase*) und
3. einer Reparatur (*repair*).

In der Originaläußerung gibt es einen Teil, das *Reparandum*, den der Sprecher korrigieren möchte. Unter Umständen bemerkt er dies mit einer gewissen Verzögerung, nachdem er das Reparandum schon geäußert hat. Er unterbricht die Äußerung. Es folgt die *Editierphase*, die aus einer Pause oder einer Pause gefüllt mit Interjektionen und Korrekturmarkern (*äh, ich meine* etc.) besteht. In der sich anschließenden *Reparatur* wird die Originaläußerung fortgesetzt, indem evtl. ein Teil davon wiederholt, das Reparandum durch ein *Reparans* ersetzt und so die Äußerung korrigiert wird.

Das Modell von Nakatani und Hirschberg Nakatani und Hirschberg (1993) legen ein etwas einfacheres Modell, das sie *Repair Interval Model* nennen, zugrunde, vgl. Abb. 2. Dieses Modell teilt eine Reparatur auch in die drei Teile Originaläußerung, Editierphase und Reparaturphase ein, verzichtet aber auch deren weitere Untergliederung.

Selbstkorrekturen und Monitoring Selbstkorrekturen sind aus psycholinguistischer Sicht ein sehr interessantes und häufig untersuchtes Phänomen. Nach Levelt (1983, 1989) übernimmt im menschlichen Sprachproduktionsprozeß eine Monitoring-Komponente die Aufgabe, die eigenen Äußerungen eines Sprechers zu parsen, während sie geäußert werden, und auf Übereinstim-

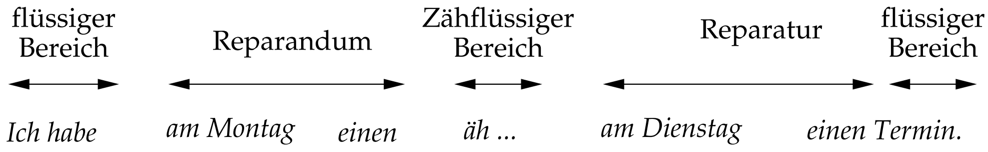


Abbildung 2: Struktur einer Korrektur nach Nakatani & Hirschberg (1993)

mung mit seinen Intentionen, auf syntaktische Korrektheit und kontextuelle Angemessenheit zu überprüfen. Er vertritt die These, diese Komponente habe keinen Zugriff auf den Sprachproduktionsprozeß, lediglich auf dessen Resultate. Stellt diese Komponente eine Abweichung oder einen Fehler fest, unterbricht sie die Sprachproduktion und setzt Reparaturprozesse in Gang.

Wohlgeformtheit von Selbstkorrekturen Für Selbstkorrekturen postuliert Levelt (1983, S. 78) die folgende Wohlgeformtheitsregel:

(1) Eine Selbstkorrektur der Form $\langle \alpha\gamma \rangle$ (α sei die Originaläußerung, γ die Reparaturphase) ist wohlgeformt genau dann, wenn es eine Wortkette β gibt, die die Konstituente, die das letzte Wort von α direkt dominiert, vervollständigt, so daß $\langle \alpha\beta$ *und* $\gamma \rangle$ syntaktisch wohlgeformt ist. Die Konjunktion *und* entfällt, wenn das erste Element von γ bereits eine Konjunktion ist.

Die Selbstkorrektur in Abbildung 1/2 ist nach dieser Regel wohlgeformt: $\alpha =$ *ich habe am Montag einen*, $\gamma =$ *einen Termin*, setze $\beta =$ *Fortbildungskurs*, dann ist $\langle \alpha\beta$ und $\gamma \rangle =$ *Ich habe am Montag einen Fortbildungskurs und einen Termin* syntaktisch wohlgeformt; das letzte Wort von α wird von einer NP dominiert (was als eine Art *top-down*-Prädiktion zu verstehen ist, da die NP nicht vollständig, ihr Auftreten aber aus dem syntaktischen Kontext vorhersagbar ist), die Vervollständigung dieser NP ist *einen Fortbildungskurs und einen Termin*.

Verarbeitung durch den Hörer Levelt (1983) formuliert zwei Regeln, mit denen der Hörer einer korrigierten Äußerung aus dieser die intendierte Äußerung ableiten könne. Die erste Regel greift, wenn das letzte Wort vor der Unterbrechungsstelle (o_n) und das erste in der Reparaturphase (r_1) von der gleichen syntaktischen Kategorie sind. Unter diesen Umständen ist o_n durch r_1 zu ersetzen und der Rest der Reparaturphase anzuhängen.

Die zweite Regel feuert, wenn es zu dem ersten Wort r_1 der Reparaturphase ein identisches o_i mit gleicher syntaktischer Kategorie in der Originaläußerung gibt. Dann ist o_i durch r_1 zu ersetzen, sowie alles was darauf folgt bis zur Unterbrechungsstelle durch den Rest der Reparaturphase.

Angewendet werden können diese Regeln allerdings nur, wenn das Vorhandensein einer Selbstkorrektur detektiert wurde. Für die maschinelle Verarbeitung stellt sich außerdem das Problem, daß die Ausgabe des Spracherkenners so deformiert sein kann, daß die relevanten Korrespondenzen zwischen den Wörtern nicht mehr hergestellt werden können.

Prosodische Markierung bei Selbstkorrekturen Laut Levelt (1989) sind bei weitem nicht alle Selbstkorrekturen prosodisch markiert. Der Auslöser einer Selbstkorrektur hat einen Einfluß darauf, ob sie prosodisch markiert wird oder nicht. Außerdem bemerkte er eine individuelle Variation zwischen den Sprechern hinsichtlich der prosodischen Markierung von Selbstkorrekturen.

Korrigiert ein Sprecher eine falsche Angabe, so ist die Wahrscheinlichkeit einer prosodischen Markierung relativ hoch: 53% dieser Korrekturen waren bei seinem Experiment durch Akzent prosodisch markiert. Von den Fällen, in denen ein Sprecher eine Formulierung durch eine angemessenere ersetzt, waren es nur 19%.

Je geringer die Zahl der möglichen Alternativen zu einer ursprünglich geäußerten falschen Angabe ist, desto häufiger ist eine prosodische Markierung. Wird z. B. eine Richtungsangabe wie *links* korrigiert, ist die Menge der Alternativen klein (vielleicht *rechts* und *geradeaus*) und die Korrektur wird prosodisch markiert durch einen Akzent auf dem Reparans.

Je größer die Alternativenmenge ist, desto schwächer fällt die prosodische Markierung aus. In seinem Experiment wurden Korrekturen von Farbadjektiven (es kamen insgesamt elf Farben vor) seltener prosodisch markiert als Korrekturen von Richtungsangaben.

Levelt (1989) widerspricht auch der Annahme, die prosodische Markierung einer Selbstkorrektur lasse auf deren syntaktische Struktur schließen. Er fand keinen Zusammenhang zwischen der Struktur der Korrektur (ob nach der Unterbrechung nur das Reparans geäußert, auch Teile der Originaläußerung wiederholt oder eine völlig neuen Äußerung begonnen wurde) und der prosodischen Markierung.

2.1.4 Untersuchungen auf dem Verbmobil-Corpus

Auf der Basis des Verbmobil-Corpus' spontansprachlicher Dialoge wurden mehrere Auswertungen zu spontansprachlichen Phänomenen durchgeführt.[1]

[1] Diese Auswertungen beziehen sich alle auf die deutschen Dialoge im Corpus. Für das Englische und Japanische liegen noch keine Untersuchungen vor.

- Tropf (1994) untersuchte 50 Dialoge mit insgesamt 641 Turns auf syntaktische Besonderheiten. Er kommt zu dem Ergebnis, daß 70% der Turns aus mehreren Sätzen oder Sätzen und freien Phrasen bestehen. In 6,4% der betrachteten Turns kommen nach dieser Studie Neuansätze (also Abbrüche) vor; nach den angegebenen Beispielen zu urteilen überlappt diese Kategorie allerdings mit der Klasse der Selbstkorrekturen, die nicht gesondert betrachtet wird. Im Vergleich mit den Ergebnissen der anderen Untersuchungen zu spontansprachlichen Phänomenen im Verbmobil-Corpus erscheint dieser Wert aber als zu niedrig.

- Tillmann und Tischer (1995) betrachteten 3873 Turns, in denen sie 1049 Fälle von Selbstkorrekturen und Abbrüchen fanden. Von diesen 1049 Fällen sind allerdings 62% sog. verdeckte Selbstkorrekturen, d. h. Häsitationen ohne echte, an der Oberfläche erkennbare Korrektur, und Wortwiederholungen. Bei dem Rest machen Abbrüche 14% aus, Korrekturen, bei denen ein Wort durch ein anderes ersetzt wird, 14%, solche, bei denen ein Wort eingefügt wird, 3%, Korrekturen, bei denen ein Wort gelöscht wird, 1%. Sonstige Anakoluthe kommen in 2% der Fälle vor.

- Tischer (1997) untersuchte 5263 Turns aus dem Verbmobil-Corpus und 3470 Turns aus freien Unterhaltungen. In diesen 8733 Turns fand er 4300 Fälle von Selbstkorrekturen und Abbrüchen, von denen verdeckte Korrekturen (Häsitationen) 30% und Wortwiederholungen 32% ausmachen. In 14% der Turns wurde eine Äußerung abgebrochen, einzelne Wörter wurden in 14% ersetzt, ein Wort oder mehrere Wörter wurden in 3% eingefügt und in 1% der Fälle getilgt. 2% machen Pivot-Konstruktionen aus, bei denen zwei Sätze miteinander über ein gemeinsames Element verbunden sind. 4% der Fälle lassen sich nicht klassifizieren.

  Anhand der transliterierten Daten versucht Tischer (1997), Regeln für die Transformation einer Äußerung, die spontansprachliche Phänomene enthält, in eine grammatische Äußerung zu formulieren. Er kommt zu dem Ergebnis, daß sich 96% der Fälle in den betrachteten Daten durch Regeln, die ausschließlich den Wortlaut und die syntaktische Struktur der Äußerung berücksichtigen, erkennen und korrigieren lassen.

  Für Selbstkorrekturen modifiziert er die oben erwähnten Regeln von Levelt (1983); für Abbrüche gibt er eine Reihe von Regeln an, die sich auf eine partielle syntaktische Analyse stützen und insbesondere auch Valenzinformation zur Detektion von Abbrüchen heranziehen. Er gibt allerdings zu bedenken, daß die Regeln nur auf Turns, die Korrekturen oder Abbrüche enthalten, getestet wurden (vermutlich manuell), also u. U. übergenerieren, wenn man sie auf Eingaben ohne spontansprachliche Phänomene anwendet. Ob sich die oberflächenorientierten Regeln auch auf verrauschte Spracherkennerausgaben erfolgreich anwenden lassen, bleibt offen.

- Finkler (1997, Kapitel 3) untersuchte 336 Dialoge und 4590 Turns, in denen er 1251 Selbstkorrekturen fand. Er betrachtete nur Wortersetzungen, d. h. keine Korrekturen, die durch syntaktische Performanzphänomene, Versprecher oder Stotterer ausgelöst werden, und keine Abbrüche.

  Nach seinen Ergebnissen sind in fast allen Fällen die syntaktische Kategorie von Reparandum und Reparans identisch. Die Unterbrechung der Originaläußerung erfolgt bevorzugt im Reparandum oder kurz danach; zwischen dem Ende des Reparandums und der Unterbrechungsstelle werden in den meisten Fällen keine weiteren Silben oder Wörter geäußert. Erfolgt ein Abbruch innhalb eines Wortes, so handelt es sich in der Regel um ein Inhaltswort.

  Der Beginn der Reparaturphase fällt gewöhnlich mit einem Phrasenbeginn zusammen. Am häufigsten treten direkte Ersetzungen (ohne Wiederholung von Material aus der Originaläußerung) auf; nur wenn das Reparandum ein Nomen ist, ist eine antizipatorische Wiederaufnahme, d. h. hier Wiederholung des Artikels oder Adjektivs vor dem Reparandum, häufiger. Die von Levelt (1983) formulierte Wohlgeformtheitsbedingung für Selbstkorrekturen fand Finkler (1997) durch die Daten des Verbmobil-Corpus' bestätigt.

- Eine eigene Auswertung der Transliteration von 637 Dialogen mit 12052 Turns nach der Häufigkeit der annotierten Phänomene Abbruch und Selbstkorrektur ergab, daß knapp 20% der Turns Selbstkorrekturen und rund 3% Abbrüche enthalten.

  Bei 95% der Selbstkorrekturen sind die korrigierten Äußerungsteile nicht länger als drei Wörter. Den größten Anteil haben Selbstkorrekturen, bei denen ein abgebrochenes Wort korrigiert wird. Den zweitgrößten Anteil haben KorrekturuKorrektururen einzelner vollständig geäußerter Wörter. Bei Selbstkorrekturen, die länger als ein Wort sind, endet die Originaläußerung in der Mehrzahl der Fälle nicht mit einem Wortabbruch.

  Abbrüche sind im Mittel vier bis fünf Wörter lang. Wortabbrüche kommen am Ende der (Phrasen-)Abbrüche kaum vor. In Abb. 3 ist die Auswertung nach der Länge für Selbstkorrekturen und in Abb. 4 für Abbrüche wiedergegeben. Dabei wird unterschieden, ob das jeweils letzte Wort im korrigierten Äußerungsteil bzw. das letzte Wort der abgebrochenen Äußerung nicht vollständig geäußert wurde.[2]

[2] Bei Wortabbrüchen hängt es von der Länge des Fragments ab, ob der Spracherkenner den Wortabbruch quasi als Störgeräusch behandelt, ein dem Fragment ähnliches oder das intendierte Wort erkennt.

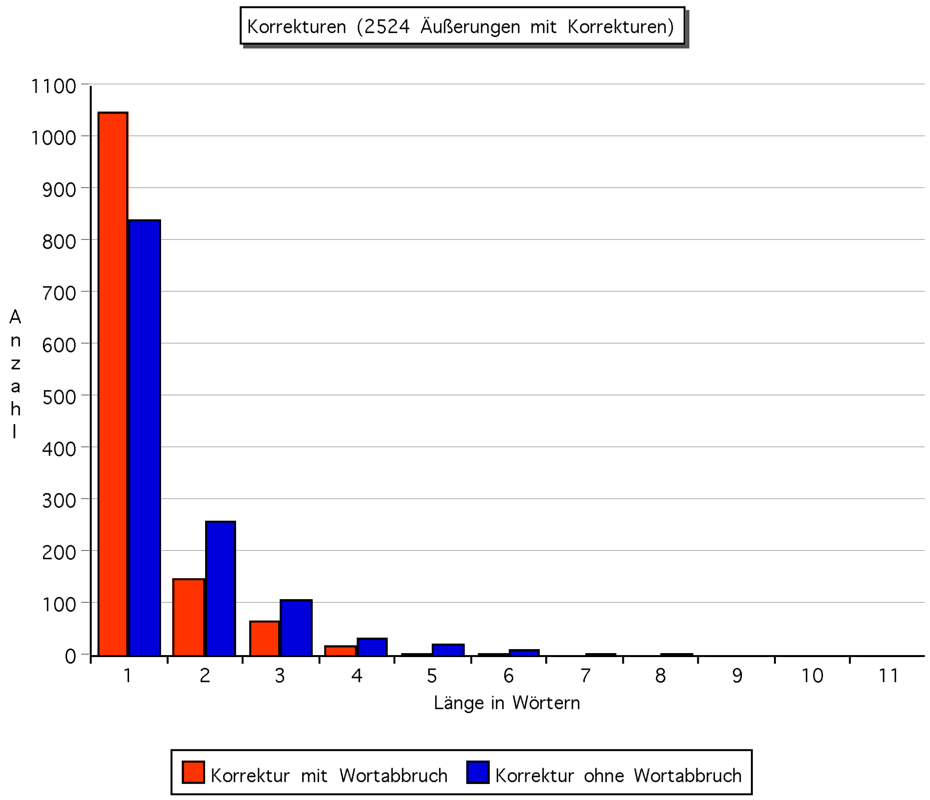


Abbildung 3: Länge der Reparanda bei Korrekturen mit und ohne Wortabbruch

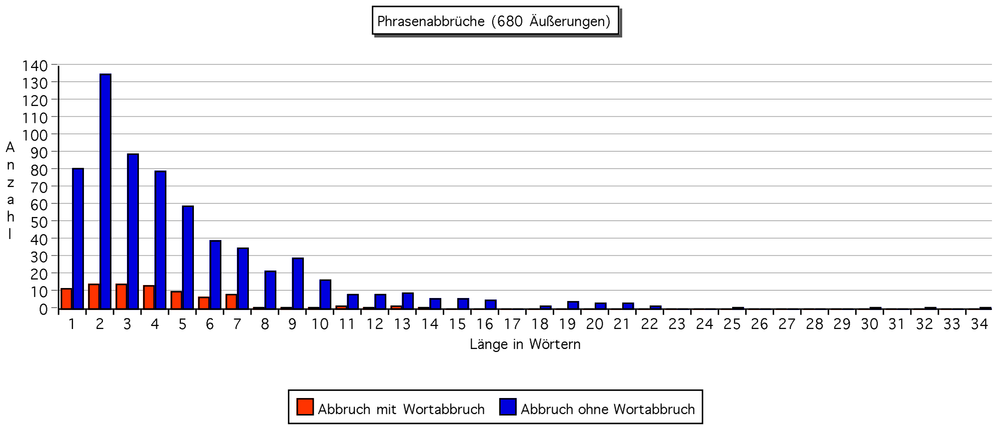


Abbildung 4: Länge der Abbrüche mit und ohne Wortabbruch

2.2 Erkennerfehler

Die Abbildung eines akustischen Sprachsignals auf eine Kette von Wörtern durch einen Spracherkenner ist nicht immer perfekt: Die errechnete Wortkette, der der Spracherkenner die höchste Wahrscheinlichkeit zubilligt, entspricht nicht in jedem Fall der tatsächlich geäußerten und ist auch häufig nicht syntaktisch wohlgeformt und analysierbar.

Aus diesem Grund liefern Spracherkenner nicht nur eine Wortkette, sondern mehrere Hypothesen über die Äußerung. Diese Hypothesen sind anhand ihrer jeweiligen Wahrscheinlichkeit, der tatsächlichen Äußerung zu entsprechen, geordnet.

Zur effizienten Repräsentation mehrerer solcher Hypothesen stellt man sie als einen gewichteten, gerichteten azyklischen Graphen dar, der als Worthypothesengraph oder Wörtergitter bezeichnet wird (ein Beispiel ist in Abb. 5 wiedergegeben) (Oerder und Ney 1993). Eine Folge von Kanten vom Anfangs- bis zum Endknoten des Graphen entspricht einer Hypothese über die Äußerung. An den Kanten sind die Wörter annotiert, die geäußert worden sein könnten, zusammen mit einer Wahrscheinlichkeit.[3]

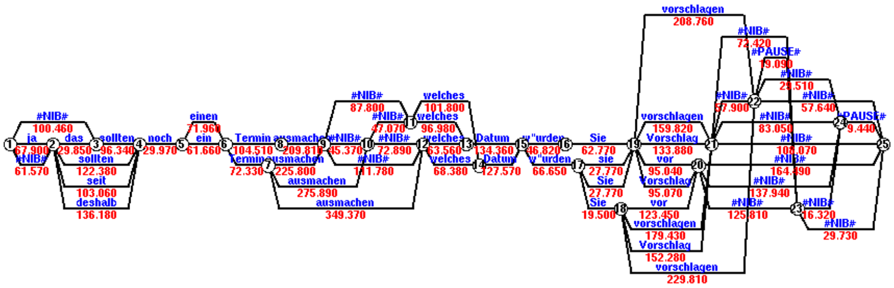


Abbildung 5: Ein Worthypothesengraph

Bei der Evaluierung der Leistung von Spracherkennern unterscheidet man drei Arten von Fehlern, die bei der Abbildung eines akustischen Sprachsignals auf eine Wortkette auftreten können:

Einfügungen: Ein nicht geäußertes Wort wird erkannt;

Auslassungen: Ein geäußertes Wort wird nicht erkannt,

Verwechslungen: Ein anderes als das eigentlich geäußerte Wort wird erkannt.

[3]Nöth und Plannerer (1993) definieren das Format der Worthypothesengraphen, das in Verbmobil verwendet wird.

Durch Vergleich der erkannten Wortkette mit der geäußerten berechnet man die Wortfehlerrate W_{err} wie folgt (E sei die Zahl der Einfügungen, A die der Auslassungen, V die der Verwechslungen; R ist die Zahl der tatsächlich geäußerten Wörter):

(2) $$W_{err} = \frac{E + A + V}{R}$$

Betrachtet man Wörtergitter, so wird für die Wortfehlerrate des Wörtergitters der Pfad durch das Wörtergitter gewertet, dessen Wortfehlerrate minimal ist (Lehning 1996), unabhängig von der akustischen Bewertung dieses Pfades. Die Satzerkennungsrate ergibt sich als Verhältnis der Zahl der vollständig korrekt erkannten Sätze zur Zahl der betrachteten Referenzsätze.

Die Spracherkenner, die zur Zeit im Verbmobil-System verwendet werden, erreichen Wortfehlerraten um 12% und Satzerkennungsraten um 20 bis 30% (Reinecke 1996). Dabei ist zu berücksichtigen, daß die Worterkennungsrate von der Dichte des Wörtergitters, definiert als das Verhältnis der Zahl der Worthypothesen zu der Zahl der Referenzwörter, abhängt: Mit steigender Dichte steigt auch die Worterkennungsrate, aber entsprechend auch die Zahl der Pfade im Wörtergitter und damit der Suchaufwand für den Parser.

Auf dem akustisch bestbewerteten Pfad wird gegenwärtig eine Wortfehlerrate von 15 bis 30% erreicht, d. h. der am wenigsten von der tatsächlichen Äußerung abweichende Pfad ist oftmals nicht der, der vom Spracherkenner am besten bewertet wird.[4]

Die linguistische Verarbeitung kann also in den meisten Fällen nicht damit rechnen, genau die tatsächlich geäußerte Wortfolge im Wörtergitter vorzufinden. Teilweise enthalten die Wörtergitter von der tatsächlichen Äußerung abweichende, aber syntaktisch analysierbare Pfade, oder sie enthalten keinen Pfad, der vollständig analysierbar ist. Leider gibt es bislang keine Statistik über die relative Häufigkeit dieser Fälle. Allerdings deuten die Testergebnisse des Verbmobil-Forschungsprototypen darauf hin, daß in vielen Fällen kein vollständig analysierbarer Pfad vorhanden ist.[5]

[4]Die Wortfehlerrate und die Satzerkennungsrate sind nicht unbedingt die einzig sinnvollen Maßstäbe, wenn es darum geht, die Leistung eines gesamten Systems zu beurteilen. Boros et al. (1996) schlagen als weiteres Maß die sogenannte Konzeptakkuratheit vor. Die Konzeptakkuratheit wird berechnet wie die Worterkennungsrate, nur werden Einfügungen, Auslassungen und Verwechslungen von semantischen Einheiten, in der von ihnen diskutierten Anwendung sind dies Merkmal-Wert-Paare, aus denen Datenbankanfragen generiert werden, betrachtet. Problematisch ist natürlich die Definition dieser semantischen Einheiten, zumal sie anwendungs- und systemabhängig sind.

[5]Die tiefe syntaktisch-semantische Analyse im Forschungsprototyp 1.0 schlägt in einem Drittel bis zur Hälfte aller Fälle fehl (vgl. die Testergebnisse der Systemgruppe). Die Ursachen sind das Nichtvorhandensein analysierbarer Pfade, das Überschreiten der Zeitbeschränkung des Parsers (was ja bedeutet, daß die akustisch gut bewerteten und daher zuerst verfolgten

Für die zweite Phase von Verbmobil ist vorgesehen, daß die Eingabe außer über Nahbesprechungsmikrophone auch über Mobiltelephone möglich sein soll. Die akustische Qualität des Signals ist auf diesem Weg schlechter, so daß mit einer höheren Wortfehlerrate zu rechnen ist. Auch durch die Ausweitung des Vokabulars wird die Wortfehlerrate zunächst wahrscheinlich steigen.

2.3 Segmentierungsprobleme

Die Segmentierung einer Äußerung in analysierbare Einheiten ist ein zentrales Problem für ein sprachverarbeitendes System, das spontansprachliche Eingabe zuläßt, da gesprochene Sprache keine Interpunktion kennt und gesprochene Äußerungen nur selten genau aus einem vollständigen Satz bestehen. Der Regelfall ist eine Aneinanderreihung von Sätzen, Satzfragmenten und einzelnen Phrasen oder Wörtern (Tropf 1994).

Eine gesprochene Eingabe wird auf verschiedenen Ebenen im Verbmobil-System segmentiert. Die prosodische Analyse ermittelt zunächst mögliche Positionen für Phrasen- und Satzgrenzen. Die Berücksichtigung dieser Grenzen durch den Parser beschleunigt das Parsing erheblich, da zahlreiche syntaktisch zulässige Segmentierungen von vornherein ausgeschlossen werden können (Kompe et al. 1997).

Die syntaktisch-semantische Analyse kann die von der Prosodie vorgegebene Segmentierung verfeinern, wenn sie einem durch zwei prosodische Grenzen definierten Segment keine überspannende Analyse zuordnen kann. In diesem Fall erzeugt sie eine Folge von Segmenten, die gemeinsam das ursprüngliche Intervall abdecken.

Eine gegebene Segmentierung einer Eingabe ist problematisch, wenn sie zu einer inhaltlich falschen Analyse führt, etwa weil Teile der Eingabe, die nicht zusammenghören, in einem Segment zusammengefaßt werden oder wenn sie zu feinkörnig ist, um die Segmente sinnvoll weiterverarbeiten zu können. Z. B. ergibt die Übersetzung einer unzusammenhängenden Sequenz von Phrasen nicht unbedingt einen kohärenter Satz.

Eine problematische Segmentierung der Eingabe kann verschiedenene Ursachen haben. So kann eine prosodische Grenze an einer Stelle erkannt worden sein, wo die Grammatik keine syntaktische Grenze zuläßt. Oder eine prosodische Grenze korrespondiert zwar zu einer syntaktischen, segmentiert die Äußerung aber dennoch zu fein, etwa weil der Sprecher weitere Äußerungsteile mit Verzögerung geäußert hat (*Ich habe Zeit . . . am Montag.)*.

Auch wenn die Eingabe wegen Erkennerfehlern nicht überspannend analysierbar ist, ist das Resultat eine Folge von relativ kleinen analysierbaren Segmenten. Eine weitere Ursache können auch Lücken in der Abdeckung der

Pfade nicht analysierbar sind), und Grammatikfehler in Syntax und Semantik.

Grammatik sein, z. B. die Nichtbehandlung topikalisierter Modifikatoren, was auch zu einer inhaltlich inadäquaten Segmentierung führt (s. 2.4).

2.4 Abdeckung der Grammatik

Die in natürlichsprachlichen Systemen eingesetzten Grammatiken sind in ihrer Abdeckung beschränkt. In einer realistischen Anwendung können Konstruktionen auftreten, die von der Grammatik nicht erfaßt werden, obwohl sie grammatisch sind. Der Wechsel zu einer anderen Domäne bringt gewöhnlich solche neuen Konstruktionen mit sich.

Aber auch innerhalb einer Domäne können einzelne Konstruktionen Probleme bereiten. Z. B. hat sich in der ersten Phase von Verbmobil die Syntax von Zeitausdrücken als ein sehr komplexes Problem erwiesen. So konnte die spezielle Zeitgrammatik des SynSemS3-Analysemoduls[6] auch in ihrer endgültigen Version nicht sämtliche Zeitausdrücke im Verbmobil-Corpus analysieren.

Neben solchen unbeachsichtigten Abdeckungslücken gibt es auch intendierte Beschränkungen der Grammatik. Um Ambiguitäten bei der syntaktischen Analyse zu reduzieren und das Parsing zu beschleunigen, werden mitunter einzelne Regeln, die zur Analyse grammatischer, aber seltener Konstruktionen erforderlich wären, nicht in die Grammatik aufgenommen.

Generell sollte eine Grammatik, die zum Parsen gesprochener Sprache, insbesondere auf Wörtergittern, eingesetzt wird, eher restriktiv sein, um den Suchraum zu beschränken. Aus diesem Grund ist auch eine Grammatik, die übergeneriert, beim Parsen gesprochener Sprache problematisch.

[6]SynSemS3 ist das gemeinsam von der Siemens AG, der Universität des Saarlandes und der Universität Stuttgart entwickelte syntaktisch-semantische Analysemodul im Verbmobil-Forschungsprototyp 1.0 (Bos et al. 1996; Block 1997).

3 Ansätze zur robusten Verarbeitung gesprochener Sprache

Es gibt eine Reihe von Ansätzen, die sich mit den Problemen der robusten Verarbeitung gesprochener Sprache befassen. Die wichtigsten werden in diesem Abschnitt kurz zusammengefaßt. In diesen Arbeiten wurden gänzlich unterschiedliche Herangehensweisen gewählt. Diese Ansätze operieren entweder

vor dem Parsing: Sie versuchen, bereits vor der syntaktischen Analyse z. B. durch spontansprachliche Phänomene verursachte Ungrammatikalität zu erkennen und zu beheben;

während des Parsings: Die Mehrzahl der Arbeiten geht von einem modizierten Parsing-Verfahren aus, bei dem etwa Teile der Eingabe ignoriert oder nur partielle Analysen geliefert werden; oder

nach dem Parsing: Diese Ansätze versuchen, aus den bruchstückhaften Resultaten der Analyse Strukturen aufzubauen, die sinnvoll weiterverarbeitet werden können.

Die meisten Arbeiten konzentrieren sich auf spontansprachliche Phänomene, gehen dabei aber zum Teil von transliterierten Daten aus. Das ist eine idealisierende Vereinfachung angesichts der tatsächlichen Ausgaben eines Spracherkenners. Einige Arbeiten erwähnen auch explizit das Problem der Erkennerfehler, ohne allerdings die dadurch verursachten Schwierigkeiten genauer zu analysieren oder spezifische Lösungen dafür vorzuschlagen.

Dieser Abschnitt gibt eine Übersicht der wichtigsten Ansätze in der Literatur; eine tabellarische Aufstellung der Merkmale der referierten Ansätze findet sich in Abb. 6, S. 23. In Abschnitt 4.6 werden die Unterschiede zwischen den Arbeiten in der Literatur und dem hier vorgeschlagenen Modell herausgearbeitet. Arbeiten zur robusten Verarbeitung natürlicher Sprache, die sich mit geschriebener Eingabe auseinandersetzen, werden nicht referiert; einen Überblick dazu gibt Stede (1992).

3.1 Vor dem Parsing

3.1.1 Erkennung spontansprachlicher Phänomene durch Pattern Matching

Bear, Dowding, und Shriberg (1992) untersuchen die Erkennung von Selbstkorrekturen in transliterierten gesprochenen Daten mittels *pattern matching* und unter Verwendung syntaktischer, semantischer und prosodischer Information. Wenn durch *pattern matching* eine mögliche Korrektur gefunden wurde, wird der korrigierte String geparst, um die Annahme zu überprüfen. Außerdem zeigen sie, daß Korrekturen z. T. anhand von Änderungen der Grundfrequenz und an der Länge der Pausen zwischen Wörtern erkannt werden können.

Heeman und Allen (1994) beschreiben ebenfalls einen Ansatz, der Selbstkorrekturen mittels *pattern matching* vor dem Parsing erkennt und korrigiert. Dabei bleibt aber unklar, wie dieser auf transliterierten Daten operierende Ansatz auf ein System zur Verarbeitung gesprochener Sprache übertragen werden soll, in dem ein Parser in der Menge der möglichen Äußerungen erst noch einen analysierbaren Pfad sucht und die Eingabe u. U. durch Erkennerfehler zu verrauscht ist, um die Muster spontansprachlicher Phänomene darin zu finden.

3.1.2 Detektion anhand von Akustik und Prosodie

Nakatani und Hirschberg (1993) untersuchen auch die Erkennung von Selbstkorrekturen mittels prosodischer Merkmale, allerdings setzen sie bereits vor Spracherkennung und Parsing bei dem akustischen Sprachsignal an. Diese prosodischen Merkmale kombinieren sie mit *pattern matching*-Techniken, um Selbstkorrekturen in transliterierten Daten zu erkennen und zu reparieren.

3.1.3 Behandlung von Selbstkorrekturen im Sprachmodell

Später untersuchen Heeman und Allen (1997) die Integration von Selbstkorrekturen in ein stochastisches Sprachmodell, wie es von einem Spracherkenner verwendet wird, das die erkannten Wörter mit *part of speech tags* annotiert und auch auf Selbstkorrekturen trainiert wurde. Ermittelt das Sprachmodell eine hohe Wahrscheinlichkeit für eine Korrektur, können die entsprechenden Teile der erkannten Äußerungen beim Parsing ausgelassen werden.

3.2 Während des Parsings

3.2.1 Parsing unter Annahme eines Korrektursignals

Hindle (1983) hat in einer mitlerweile klassischen Arbeit die hohe Regularität von Selbstkorrekturen gezeigt. Sein Ansatz leidet aber unter der Annahme eines phonetischen Korrektursignals an der Unterbrechungsstelle, das in den Transliterationen, die er zugrundelegt, bei Auftreten einer Selbstkorrektur annotiert wurde. Er modifiziert einen Parser so, daß er bei Auftreten eines Korrektursignals in der Eingabe überprüft, ob vor und hinter dem Signal Konstituenten gleicher Kategorie auftreten. Wenn dies der Fall ist, oder vor dem Signal ein unvollständiges Präfix einer Konstituente der gleichen Kategorie wie nach dem Signal steht, wird die Konstituente bzw. das Präfix vor dem Signal gelöscht.

3.2.2 Case-frame Parsing

Hayes et al. (1986) beschreiben den vermutlich ersten implementierten Versuch, von einem Spracherkenner erkannte gesprochene Eingaben robust zu parsen. Sie verwenden *case-frame parsing*, um von den Ungrammatikalitäten spontansprachlicher Eingaben und der durch Erkennerfehler bedingten Nichtwohlgeformtheit zu abstrahieren. Inhaltswörter wie Verben und Nomina sind als *case-frame headers* ausgezeichnet, nach denen der Parser zunächst sucht. Jeder verbale *frame* hat eine Menge zugeordneter *slots*; der Parser füllt diese mit nominalen *frames*. Dieses Vorgehen setzt voraus, daß die Domäne sehr stark eingeschränkt ist, da sonst die Kombinationsmöglichkeiten der *frames* ohne Beschränkung durch syntaktische Information explodieren.

3.2.3 Iteriertes Parsen

Langer (1990) hat ein Verfahren zum Parsing von Äußerungen mit Selbstkorrekturen vorgeschlagen, daß bei Auftreten von korrekturmarkierenden Interjektionen (z. B. *halt, nein*) sukzessive Wörter vor der Interjektion und diese selbst entfernt, bis die resultierende Zeichenkette parsbar ist. Fälle von Korrekturen, bei denen keine solche Interjektion vorhanden ist, können damit nicht behandelt werden. Außerdem ist das wiederholte Parsing ineffizient.

3.2.4 Partielles Parsing

Eckert und Niemann (1994) argumentieren, durch spontansprachliche Phänomene und Erkennerfehler deformierte Eingaben seien nur durch partielle Analyse zu bewältigen. Ihr *bottom-up* operierender Chartparser produziert eine Liste partieller Analyseresultate, die anhand ihrer Abdeckung der Eingabe, ihrer akustischen Bewertung und ihrer syntaktischen und semantischen Wohlgeformtheit bewertet werden. Im Falle einer Selbstkorrektur, bei der eine PP durch eine weitere PP korrigiert wird, liefert ihr Parser zwei partielle Analysen für die PPen, von denen nur die zweite weiterverarbeitet wird. Die partiellen Analyseresultate sind die Eingabe für eine anwendungsspezifische Komponente, die daraus Datenbankanfragen generiert (es handelt sich um ein Fahrplanauskunftssystem).

3.2.5 Parsing mit kontextuellen Vorhersagen

Hanrieder und Görz (1995) arbeiten ebenfalls mit partiellen Analysen, wenn keine vollständige zu erzielen ist. Bei ihrem Ansatz sucht ein Inselparser ausgehend von akustisch hochbewerteten Kanten im Wörtergitter nach analysierbaren (Teil-)Pfaden (Ehrlich und Hanrieder 1996). Zusätzlich beziehen sie Vorhersagen über Sprecheräußerungen in die Bewertung der partiellen Analysen ein: Der Score einer partiellen Analyse hängt ab von ihrer akustischen Bewertung, ihrer syntaktischen Vollständigkeit und dem Grad ihrer Übereinstimmung mit den Vorhersagen über mögliche Äußerungen. Die ausgewählten partiellen Analysen werden an eine Komponente weitergegeben, die Anfragen an eine Fahrplandatenbank erzeugt.

3.2.6 Minimal Distance Parsing

Hipp (1993) überträgt das *minimal distance parsing* auf Wörtergitter. Wenn der Parser beim *minimal distance parsing* eine Eingabe nicht analysieren kann, versucht er, die Eingabe durch Löschungen und Einfügungen von Wörtern soweit zu verändern, daß sie parsbar wird. Diese Änderungen sind minimal in dem Sinne, daß durch die Änderungen ein modifizierter String mit möglichst geringer Abweichung von der ursprünglichen Eingabe erzeugt wird. Die Abweichung wird analog zur Wortfehlerrate eines Spracherkenners berechnet (vgl. 2.2), wobei die Lösch- und Einfügeoperationen abhängig vom Wort gewichtet werden. Da im allgemeinen mehrere modifizierte Strings eine (fast) gleiche Abweichung zur Eingabe haben, werden die resultierenden semantischen Repräsentationen mit den Vorhersagen einer Dialogkomponente verglichen.

Ob *minimal distance parsing* in Anwendungen realistischer Größe anwendbar wäre, ist fraglich. Das System von Hipp (1993) hat ein Vokabular von nur 125 Wörtern. Die angegebenen Ergebnisse beziehen sich auf Testdaten, die zur Hälfte aus trivialen Äußerungen (ein Wort) bestehen. Außerdem wird eine Analyse als korrekt gewertet, wenn die Hälfte der Bedeutung korrekt berechnet wurde, ohne daß dieses Kriterium definiert wird. Die Komplexität des Parsingalgorithmus ist $O(V^3k^2)$ (V ist die Zahl der Knoten im Wörtergitter, k die Zahl der Regeln der Grammatik). (Vgl. die Ausführungen von Rosé (1997) und Rosé und Lavie (1997) zur Kritik an diesem Ansatz.)

3.2.7 Parsen mit Lücken

Lavie (1996) entwickelt einen Parser für gesprochene Sprache, der bei einer nicht vollständig parsbaren Eingabe versucht, durch Ignorieren von Wörtern den maximalen analysierbaren Teilstring zu finden. Dieser Parser ist auch die Grundlage der Arbeiten von Rosé (1997) (s. 3.3) und wurde in das JANUS-System (Lavie et al. 1996) zur maschinellen Übersetzung gesprochener Sprache integriert.

3.2.8 Parsing ohne Syntax

Antoine (1996) beschreibt ein Parsingverfahren für gesprochene Sprache, das ohne Rückgriff auf eine syntaktische Analyse auskommt. Es basiert auf der sog. *Microsemantics*, einer lexikalisch-semantischen Spezifikation der möglichen Argumente eines Lexems und ihrer sortalen Eigenschaften. Durch semantisches *priming* werden *top-down*-Vorhersagen an den Spracherkenner weitergegeben, um die Perplexität zu reduzieren. Das eigentliche Parsing besteht in der Unifikation der mikrosemantischen Argumentrahmen.

3.2.9 Dialogaktparsing

Quantz und Alexandersson (1996) schlagen ein Parsing-Verfahren vor, bei dem nicht eine klassische syntaktisch-semantische Analyse das Ziel ist. Ihr Parser liefert stattdessen eine Folge von Dialogakten, die den zentralen Gehalt der geparsten Äußerung wiedergeben. Die Dialogakte sind zum Teil domänenspezifisch, zum Teil aber auch domänenunabhängig. Quantz und Alexandersson (1996) argumentieren, Dialogakte seien die richtige Abstraktionsstufe für die maschinelle Übersetzung.

Der Parser ist implementiert als ein Kellerautomat, der im Wörtergitter nach Phrasen sucht und diese miteinander kombiniert. Syntaktische Restriktionen wie Kongruenz werden dabei ignoriert, aber bei der Bewertung der Parsingergebnisse herangezogen, ebenso wie die akustische Bewertung des zugrundeliegenden Pfades und die syntaktisch-semantische Vollständigkeit in bezug auf (un-)gebundene Argumente.

3.2.10 Hybride Verarbeitung

Rayner und Carter (1997) setzen im *Spoken Language Translator* (Carter et al. 1997) sowohl regelbasierte als auch stochastische Methoden ein, um die Robustheit des Systems zu erhöhen. Im ersten Analyseschritt wird die Eingabe lexikalisch und morphologisch analysiert. Dann werden mittel eines stochastischen Modells unwahrscheinliche Hypothesen über die Kategorien von Wörtern entfernt. Im nächsten Schritt wird eine Menge von Syntaxregeln angewendet, die Phrasen aufbaut. Nach diesem Schritt werden nun unwahrscheinliche phrasale Hypothesen entfernt. Schließlich wird eine Menge von mittels *explanation based learning* (Samuelsson 1994) aus der ursprünglichen Grammatik abgeleiteter Regeln angewendet, die Phrasen zu größeren Einheiten kombinieren.

Zur Erhöhung der Robustheit als Übersetzungssystem wird nach jedem Analyseschritt eine Übersetzung erzeugt, schon nach der lexikalisch-morphologischen Analyse durch Wort-für-Wort-Ersetzungen, und durch ein semantisches Transfermodul nach der phrasalen Analyse und schließlich nach dem abschließenden Aufbau einer vollständigen Analyse. Die Übersetzungsergebnisse der einzelnen Schritte werden, ähnlich wie bei Frederking und Nirenburg (1994), in eine *chart* eingetragen, aus der damit zu jeder Zeit Übersetzungen extrahiert werden können.

3.3 Nach dem Parsing

3.3.1 Verarbeitung von Interlingua-Fragmenten

Dem hier vorgeschlagenen Ansatz am ähnlichsten ist das Modell von Rosé (die bestehenden Unterschiede werden in 4.6 diskutiert). Rosé und Waibel (1996) beschreiben ein Modul, das aus partiellen Analysen größere Strukturen aufbaut. Das Modul operiert auf semantischen Repräsentationen, genauer auf *frame*-basierten interlingualen Repräsentationen von Äußerungsbedeutungen in einem System zur maschinellen Übersetzung (JANUS), die als Merkmalsstrukturen dargestellt sind.

Das Modul versucht, ungefüllte *slots* eines *frame*s mit bisher unangebundenen Elementen zu füllen. Zur Beschränkung der Kombinatorik werden Statistiken über die Kombinationswahrscheinlichkeit interlingualer Konzepte und Rückfragen an den Benutzer verwendet. Der wesentliche Nachteil dieses Ansatzes ist laut Rosé und Waibel (1996), daß dem Benutzer zu viele spezifische Fragen gestellt werden.

In einer Fortentwicklung dieses Ansatzes erzeugt Rosé (1996) anstatt lokaler Hypothesen über die Kombinierbarkeit von einzelnen Fragmenten globale Reparaturhypothesen, die beschreiben, wie mehrere Eingabefragmente in eine Gesamtrepräsentation integriert werden können. Die Hypothesen werden bewertet und anhand ihrer Komplexität (gemessen an der Zahl der notwendigen Operationen), der erzielten Abdeckung der Eingabe und von Statistiken über die Häufigkeit von *slot-filler*-Kombinationen in eine Rangfolge gebracht. Durch Fragen an den Benutzer wird dann eine Hypothese ausgewählt, wobei die Zahl der Rückfragen in diesen Klärungsdialogen möglichst gering gehalten wird.

Der hohe Abstraktiongrad der interlingualen Repräsentationen bringt den Nachteil mit sich, daß sehr viele Kombinationen von Segmenten möglich sind. Formal wird die Kombinierbarkeit nur eingeschränkt durch die Syntax der Interlingua. Rosé (1996) diskutiert das Beispiel *Thursday I am out*, das ihr Parser in zwei Segmente spaltet, da die Grammatik topikalisierte Modifikatoren nicht zuläßt. *I am out* wird abgebildet auf einen *frame* vom Typ *busy*, worin *Thursday* alternativ die *slots when* (temporale Lokalisierung) oder *why* (Begründung) füllen kann.[7] Bei Einbeziehung linguistischer Information wird der Suchraum deutlich kleiner (das Beispiel wird in 5.3 diskutiert).

3.3.2 Verarbeitung syntaktischer Fragmente

Nederhoff et al. (1997) beschreiben die syntaktische und semantische Analyse in einem Dialogsystem zur Erteilung von Fahrplanauskünften. Wenn ihr

[7] Nach der vollständigen Definition der Interlingua (Rosé 1997) kommen sogar sechs *slots* in Frage.

Parser keine Analyse findet, die die gesamte Eingabeäußerung überspannt, sucht er entlang der Pfade im Worthypothesengraphen interpretierbare Konstituenten (Phrasen) und kombiniert sie. Die Auswahl eines Pfades erfolgt dabei anhand der akustischen Bewertung und der Länge der zwischen den in die Analyse einbezogenen Konstituenten liegenden Lücken. Der Kombinationsmechanismus wird nicht näher beschrieben.

Auch Stallard und Bobrow (1992) entwickeln einen *fallback*-Mechanismus, der syntaktische Analysefragmente kombiniert, wenn keine überspannende Analyse gefunden wird. Zunächst werden spezifische Grammatikregeln angewendet, die Konstituenten kombinieren, obwohl nichtinterpretierte Äußerungsteile dazwischen stehen. Ist dies nicht erfolgreich, werden Heuristiken angewendet, die die Semantik der Fragmente anhand domänenspezifischer *slot-filler*-Kombinationen miteinander verknüpfen.

Autoren	Eingabe	Schwerpunkt	Position
Bear, Dowding, und Shriberg (1992)	TL	SP	vor
Nakatani und Hirschberg (1993)	TL	SP	vor
Heeman und Allen (1994)	TL	SP	vor
Heeman und Allen (1997)	TL	SP	vor
Hayes et al. (1986)	WHG	SP, EF	während
Hindle (1983)	TL	SP	während
Langer (1990)	TL	SP	während
Hipp (1993)	WHG	SP, EF	während
Eckert und Niemann (1994)	n-best/WHG	SP, EF	während
Hanrieder und Görz (1995)	WHG	SP, EF	während
Mayfield et al. (1995)	n-best/WHG	SP	während
Antoine (1996)	n-best/WHG?	SP	während
Quantz und Alexandersson (1996)	WHG	SP, EF	während
Lavie (1996)	n-best/WHG	SP, EF	während
Rayner und Carter (1997)	n-best/WHG?	SP, EF	während
Stallard und Bobrow (1992)	n-best/WHG?	SP	nach
Nederhoff et al. (1997)	WHG	SP, EF	nach
Rosé (1997)	n-best/WHG	SP	nach

Abbildung 6: Klassifikation der Ansätze anhand ihrer Eingabe (auf der sie getestet wurden; TL: Transliterationen, n-best-Listen, WHG: Wörtergitter), ihres Schwerpunkts (SP: spontansprachliche Phänomene, EF: Erkennerfehler) und ihrer Position in der Systemarchitektur (vor, während, nach dem Parsing)

4 Robuste semantische Verfahren

Dieser Abschnitt beschreibt die vorgesehene Funktionsweise der *robusten semantischen Verfahren* im Verbmobil-System. Die robusten semantischen Verfahren sollen als ein Modul implementiert werden, das auf den Ausgaben der unterschiedlichen linguistischen Analysekomponenten in Verbmobil 2 operiert. Dieses Modul soll Folgen von (teilweise partiellen) semantischen Repräsentationen, VITs[8] (Bos, Schiehlen, und Egg 1996), unter Berücksichtigung von Informationen aus Syntax, Semantik, Prosodie, Dialogverarbeitung und Kontextauswertung auf Folgen von VITs abbilden, wobei spontansprachliche Phänomene ausgefiltert und bruchstückhafte Analysen zu größeren Einheiten integriert werden sollen.

Zunächst wird auf die geplante Funktionalität der robusten semantischen Verfahren, die Einordnung dieses Moduls in das Gesamtsystem, das zugrundeliegende Verarbeitungsmodell und die bestehenden Anforderungen an andere Systemkomponenten eingegangen. Abschließend werden die Unterschiede zwischen dem hier vorgeschlagenen Modell und den Ansätzen in der Literatur herausgestellt.

4.1 Funktionalität der robusten semantischen Verfahren

Das Ziel der robusten semantischen Verfahren ist, Lösungen für die in Abschnitt 2 dargestellten Probleme der Verarbeitung von Spontansprache bereitzustellen. Die Kernaufgabe besteht darin, Folgen von fragmentarischen semantischen Analysen zu größeren Einheiten zusammenzufügen, wobei auch Resultate verschiedener Parser miteinander kombiniert werden. Damit wird die Qualität der Analyse und somit auch der Übersetzung erhöht:

- Wenn die Analyse fragmentarisch ist, weil die Äußerung einen Abbruch oder eine Selbstkorrektur enthält, sind die abgebrochenen oder korrigierten Anteile herauszufiltern und die übrigen Bruchstücke zu einer konsistenten Analyse zu kombinieren.
- Wenn aufgrund von Erkennerfehlern keine überspannende Analyse erreicht wurde, sind von den verschiedenen Parsern abgelieferte Fragmente so zu integrieren, daß eine Analyse entsteht, die der Bedeutung der eigentlichen Äußerung möglichst nahe kommt. Auch durch mangelnde Abdeckung der Grammatik oder durch Segmentierungsprobleme verursachte Fragmentierungen werden durch Zusammenfügen von Segmenten behandelt.

Die Qualität der Ausgabe der robusten semantischen Verfahren hängt von der Qualität der Eingaben ab. Je weniger defizient die Eingabe ist, desto besser

[8] VIT steht für *Verbmobil Interface Term*.

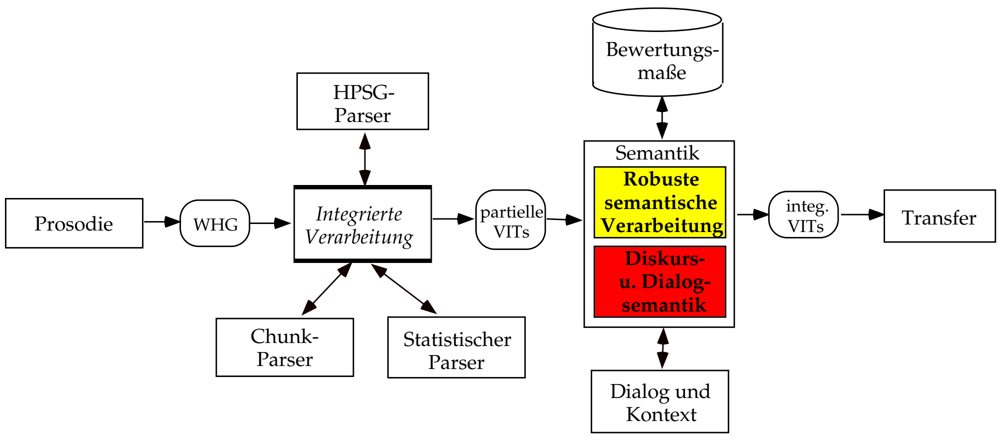


Abbildung 7: Ausschnitt aus der Systemarchitektur

ist das Ergebnis. Ist die Eingabe stark deformiert und fehlen partielle Analysen zu wichtigen bedeutungstragenden Äußerungsteilen, weil entsprechende Hypothesen nicht im Wörtergitter vorhanden sind oder von keinem Parser gefunden wurden, können die robusten semantischen Verfahren keine wesentliche Verbesserung erzielen.

Es ist aber anzunehmen, daß die Verknüpfung der verschiedenen Analyseverfahren deren jeweilige Stärken bündeln und so ihre Schwächen ausgleichen kann. Beispiele für die intendierte Funktionalität der robusten semantischen Verfahren werden in Abschnitt 5 diskutiert.

4.2 Einordnung in die Verarbeitungsabläufe

Die robusten semantischen Verfahren, die als Submodul *RoSI* (für *Robust Semantic Interpretation*) des Moduls *Semantics* implementiert werden sollen, setzen nach dem Parsing an und operieren auf den Resultaten der verschiedenen linguistischen Analyseverfahren (tiefer HPSG-Parser, Chunk-Parser, statistischer Parser), die durch das Modul *integrierte Verarbeitung* gesteuert werden, vgl. Abb. 7. Die einzelnen Parser arbeiten dabei parallel auf einem Pfad durch das Wörtergitter, den die integrierte Verarbeitung innerhalb eines durch die prosodische Analyse vorgegebenen Segments auswählt.

Schlägt die Analyse dieses Pfades fehl, wählt die integrierte Verarbeitung einen anderen aus und gibt ihn zur Analyse an die Parser. Bis eine vollständige Analyse gefunden wird oder die Zeitschranke für das Parsing erreicht ist, werden die Analyseergebnisse, VITs, an das Modul robuste semantische Verfahren weitergegeben. Diese VITs sind partiell, wenn durch die Analyse kein grammatischer (Teil-)Pfad gefunden werden konnte.

Neben der Übergabe der Resultate über die integrierte Verarbeitung an die robusten semantischen Verfahren ist nach Einführung des Pool-Konzepts[9] zur Kommunikation zwischen den Verbmobil-Modulen auch der Zugriff auf partielle Analyseresultate über den Pool möglich, wo die Ergebnisse der verschiedenen Parser abgelegt werden können. Auf die Resultate der Parser, die inkrementell Ergebnisse produzieren (z. B. der HPSG-Parser), können die robusten semantischen Verfahren dann auch inkrementell angewendet werden.

Durch diese Positionierung in der Architektur und den Zugriff auf den Pool ist es möglich, auf die Resultate der einzelnen Analysekomponenten zuzugreifen und diese untereinander zu kombinieren. Eine weitere wichtige Voraussetzung dafür ist, daß die Resultate im VIT-Format geliefert werden, das von der vom jeweiligen Parser verwendeten Grammatik und der darin enthaltenen Semantikkonstruktion unabhängig ist.[10]

Als Ausgabe liefern die robusten semantischen Verfahren wiederum VITs, an den Transfer, der sie in zielsprachliche VITs übersetzt, und an die Dialogkomponente, die die VITs im Dialoggedächtnis ablegt und protokolliert, welche Segmente zusammengefügt, modiziert oder gelöscht wurden.

4.3 Verarbeitungsmodell und Implementation

Für die Implementation der robusten semantischen Verfahren ist ein regelbasiertes Verarbeitungsmodell geplant. In den Regeln wird spezifiziert, unter welchen Bedingungen VITs zusammengefügt, modifiziert oder gelöscht werden.

Die Verarbeitung geht von den (partiellen) Analysen aus, die von den Parsern geliefert wurden. Durch Regelanwendung werden weitere (ggf. immer noch partielle) Analysen erzeugt, die größere Teile der Eingabe abdecken, bis die Eingabe vollständig abgedeckt ist oder keine Regel mehr angewendet wer-

[9]Das Pool-Konzept wurde von Johannes Schwinn in einem Vortrag auf dem Workshop zur Verbmobil 2-Architektur in Kaiserslautern (April 1997) vorgestellt. Es soll die bisherige Kommunikationsarchitektur zwischen den Modulen im Verbmobil-System ablösen, die nur von direkter Kommunikation von Modul zu Modul ausging.

Beim Pool-Konzept legt jedes Modul, das Daten produziert, diese in einem Datenpool ab. Auf den Datenpool kann jedes andere Modul zugreifen, z. B. indem es Daten eines bestimmten Typs abonniert. Wenn neue Daten dieses Typs vorliegen, wird das abonnierende Modul automatisch damit versorgt.

Mit dem Pool-Konzept wird die strikte Sender-Empfänger-Kommunikation abgelöst durch eine offene Architektur mit freiem Informationsfluß, bei der der Produzent eines Datums nicht wissen muß, wer es konsumiert, und der Konsument Daten erhalten kann, ohne mit dem Produzenten diese Kommunikation explizit zu vereinbaren.

[10]VITs enthalten sowohl semantische als auch syntaktische Information. Die semantische Information im VIT ist für alle Parser standardisiert. Lediglich bei der syntaktischen Information ist es zulässig, daß ein Parser Information, die ihm nicht zur Verfügung steht, nicht in den VIT einträgt.

den kann. Die Regeln werden also *bottom-up* bzw. vorwärtsverkettend angewendet.

Um die Regeln effizient anzuwenden und die von mehrerern Parsern gelieferten Analyseergebnisse miteinander kombinieren zu können, werden die einzelnen Analysen als Kanten in eine *chart*-artige Datenstruktur eingetragen. Auf dieser *chart* operieren die Regeln und fügen neue Kanten hinzu.

Die Kanten tragen Bewertungen, die von der akustischen Bewertung des entsprechenden Pfads im Wörtergitter, von ihrer Länge und der Abdeckung der Eingabe, von der Zahl noch ungebundener Argumente usw. abhängen. Anhand dieser Bewertungen wird eine Folge von (ggf. partiellen) Resultaten ausgewählt und ausgegeben.

Diese *chart* kann als eine Realisierung des Konzeptes des Interpretationsgraphen (Wahlster 1997) betrachtet werden. Analyseergebnisse aus verschiedenen Quellen werden untereinander verglichen und kombiniert. Selbst wenn keine überspannende Analyse erzielt werden kann, ist die Auswahl einer Folge von partiellen Ergebnissen möglich.

4.3.1 Regelformat

Die robusten semantischen Verfahren sollen regelbasiert arbeiten. Eine Regel muß spezifizieren, unter welchen Bedingungen sie anwendbar ist, welche Operationen bei ihrer Anwendung auszuführen sind, und was das Resultat ihrer Anwendung ist. Die Regeln haben daher drei Teile:

- Eine Abfolge von *Bedingungen über VITs*, in der Reihenfolge, in der die VITs in der Eingabe erwartet werden,
- eine Reihe von *Operationen* auf den Eingabe-VITs, und
- die Spezifikation eines VITs als *Ausgabe* der Regel.

Da die Implementation der robusten semantischen Verfahren in PROLOG erfolgt, werden die Regeln in PROLOG-Notation wie folgt dargestellt:

(3) `ListOfConditions ---> ListOfOperations & OutputVIT.`

`ListOfConditions` ist die Liste der Bedingungen; die Reihenfolge der Listenelemente entspricht der zeitlichen Ordnung der Eingabe-VITs, d. h. das erste Element ist eine Bedingung über den ersten Eingabe-VIT, das zweite eine Bedingung über den zweiten, usw. Mehrere Bedingungen über einen Eingabe-VIT werden in einer Liste zusammengefaßt.

Die zweite Liste, `ListOfOperations`, enthält die Operationen, die auf den Eingabe-VITs auszuführen sind. `OutputVIT` ist eine Variable über einen VIT, die entweder durch einen Eingabe-VIT oder das Resultat einer Operation gebunden sein muß, und gibt an, welcher VIT die Ausgabe der Regel ist.

```
(4)  [unbound_arg(V1,Type),type(V2,Type)] --->
     [apply(V1,V2,V3)] & V3.
```

Die Beispielregel in (4) konsumiert zwei VITs, wenn dem ersten (`V1`) noch ein Argument vom Typ `Type` fehlt und der zweite VIT (`V2`) von diesem Typ ist. Es wird dann der erste als Funktor auf den zweiten angewendet und das Resultat, ein dritter VIT (`V3`), ist die Ausgabe.

Im Gegensatz zu den vom Transfer in Verbmobil eingesetzten Regeln (Dorna und Emele 1996) operieren diese Regeln nicht auf Mengen von Prädikaten in VITs, sondern auf vollständigen VITs (wobei der Zugriff auf die Information im VIT natürlich gegeben ist). Während die Transferregeln einen quellsprachlichen VIT sukzessive abbauen und parallel einen zielsprachlichen aufbauen, kombinieren diese Regeln mehrere VITs oder modifizieren oder tilgen einzelne VITs. Eine solche Regel kann auch mehr als einen Eingabe-VIT konsumieren.

Ein Beispiel wird in Abschnitt 4.3.3 diskutiert; weitere Beispiele für Regeln und ihre Anwendung sind in Abschnitt 5 aufgeführt.

4.3.2 Regelanwendung und *chart*

Die robusten semantischen Verfahren bekommen partielle Analysen von verschiedenen Analysekomponenten, u. U. auch von einer Komponente wegen Ambiguitäten mehrere Analysen für einen Pfad oder sogar konkurrierende Analysen für mehrere Pfade im Wörtergitter. Um diese Vielfalt an Resultaten zu speichern und zu berechnen, wie einzelne davon optimal in eine Gesamtanalyse integriert werden können, ist eine Datenstruktur erforderlich, die es ermöglicht, diese Berechnungen effizient durchzuführen und Zwischenergebnisse wiederverwendbar abzulegen.

Die partiellen Analyseresultate können effizient als Einträge einer *chart*-Datenstruktur, wie sie auch beim Parsing verwendet wird (Winograd 1983; Gazdar und Mellish 1989; Wirén 1992), dargestellt werden.[11] In diese *chart* werden die partiellen Analysen mit ihrer jeweiligen Abdeckung eingetragen. Dann werden systematisch die Kombinationsregeln auf die *chart*-Einträge angewendet und neue partielle Analysen mit i. allg. größerer Abdeckung[12] berechnet.

[11]Eine *chart* als Datenstruktur für die Resultate alternativer Komponenten benutzen auch Frederking und Nirenburg (1994). Sie verwenden eine *chart*, um die Ergebnisse mehrerer Übersetzungssysteme auf einem quellsprachlichen Text untereinander zu kombinieren. Diese Kombination der Übersetzungsresultate geschieht aber auf rein textueller Ebene und durch Benutzerinteraktion, nachdem das System auf der Basis von Bewertungsmaßen eine Folge von Segmentübersetzungen ausgewählt hat, die die Eingabe abdecken.

[12]Es sind auch Regeln denkbar, die einen VIT modifizieren, ohne ihn mit anderen zu kombinieren, also zu einer Kante in der *chart* eine weitere mit gleicher Abdeckung hinzufügen. Eine mögliche Anwendung solcher Regeln ist z. B. das Abbinden eines fehlenden Arguments durch einen unterspezifizierten Platzhalter.

Der Algorithmus terminiert, wenn keine Regeln mehr anwendbar sind, oder wenn von außen eine Abfolge von Analysen angefordert wird. Dann wird eine die Äußerung abdeckende Folge von Analysen ausgewählt; es gibt also drei Phasen der Verarbeitung (vgl. Abb. 8):

Eintragen partieller Analysen: Die Eingaben von den verschiedenen Parsers werden in die *chart* eingetragen.

Regelanwendung: Durch Anwendung von Regeln werden neue Einträge in der *chart* erzeugt.

Auswahl eines Resultats: Anhand der Bewertungen wird eine Folge von Einträgen aus der *chart* ausgewählt und ausgegeben.

Diese Phasen werden u. U. mehrfach durchlaufen, etwa wenn die Ausgaben eines Parsers bereits eingetragen und die Regeln darauf angewendet worden sind und dann die Ergebnisse eines anderen Parsers eintreffen.

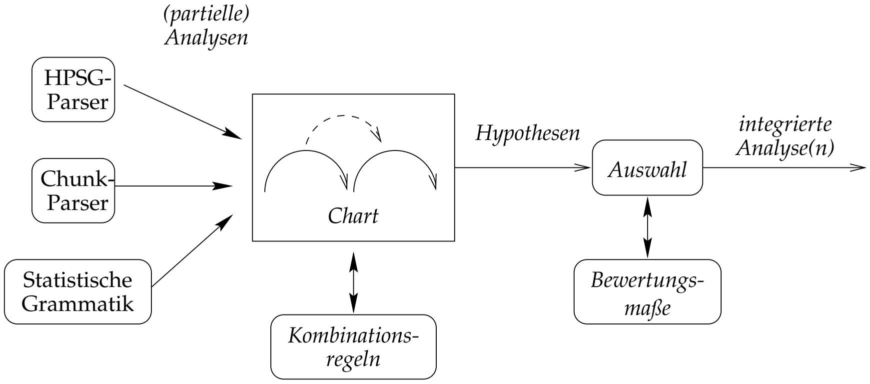


Abbildung 8: Die *chart* als zentrale Datenstruktur

Gegenüber einer klassischen *chart* sind einige Erweiterungen und Verallgemeinerungen nötig:

- Es sollen Analysen in die *chart* eintragen werden, die unterschiedlichen Pfaden im Wörtergitter entsprechen; man kann die *chart* daher nicht anhand der Stringpositionen indizieren, die eine Analyse abdeckt, sondern muß Anfangs- und Endzeitpunkt der Teiläußerung verwenden.

- Zwischen durch eine Regel kombinierbaren *chart*-Einträgen kann es kleine Lücken geben (verursacht durch Sprechpausen, Störgeräusche, etc.); bei der Regelanwendung muß es daher zulässig sein, daß zwei Einträge, die nicht unmittelbar adjazent sind, kombiniert werden.

- Wenn zeitlich aufeinanderfolgende Einträge nicht zu kombinieren sind, muß geprüft werden, ob durch Überspringen von Einträgen eine Analyse mit größere Abdeckung erreicht werden kann. Diese Möglichkeit muß aber eingeschränkt werden, da die Zahl der zu überprüfenden Möglichkeiten sonst exponentiell steigt und das Verfahren ineffizient wird.[13]

Der Vorteil einer *chart* beim Parsing und bei dem hier skizzierten Vorhaben liegt darin, daß Zwischenergebnisse, die an mehreren Stellen im Suchraum benötigt werden, nur einmal berechnet und gespeichert werden. Eine Kombination von zwei partiellen Analysen kann Teil mehrerer größerer Hypothesen sein, muß aber nur einmal berechnet werden.[14]

Aus der *chart* kann zu jeder Zeit anhand der Bewertungsfunktionen eine Folge von Hypothesen ausgewählt werden, die gemeinsam die gesamte Eingabe überspannen. Abhängig davon, wieviele Regelanwendungen bereits erfolgt sind, wird dieses Ergebnis mehr oder weniger fragmentiert sein; die Qualität steigt mit der zur Verfügung stehenden Zeit. Es handelt sich also um ein *anytime*-Verfahren (Dean und Boddy 1988; Boddy 1991), was erhebliche Vorteile für ein sprachverarbeitendes System bietet, das unter Zeitbeschränkungen arbeiten soll (Wahlster 1993).

4.3.3 Berechnungsbeispiel

Betrachten wir als Beispiel eine Selbstkorrektur, bei der der Sprecher seine Äußerung mit *das ist* … begonnen, sie dann aber mit einem anderen Verb fortgesetzt hat: … *paßt mir sehr gut*.

Angenommen, der Parser hat eine Sequenz von VITs zu dem folgenden vollständig korrekt erkannten Pfad durch das Wörtergitter geliefert:

(5) `das ist paßt mir sehr gut`

Wenn der Parser diesem Pfad drei fragmentarische Analysen zuordnet, nämlich für *das*, *ist* und *paßt mir sehr gut*, erhält man für diese Segmente insgesamt drei initiale Charteinträge.

Nun kann die Regel (6), die einen Funktor mit einem potentiellen Argument kombiniert, zweimal angewendet werden, einmal auf *das* und *ist* und zum zweiten auf *das* und *paßt mir sehr gut*.

(6)
```
[unbound_arg(V1,Type),type(V2,Type)] --->
[apply(V1,V2,V3)] & V3.
```

[13] Hier kann man es sich zu Nutze machen, daß korrigierte Äußerungsanteile in den meisten Fällen sehr kurz sind (vgl. 2.1.4) und so die maximal zulässige Länge einer Lücke beschränken.

[14] Es wird also Gebrauch von Techniken des *dynamic programming* (Cormen, Leiserson, und Rivest 1990) gemacht.

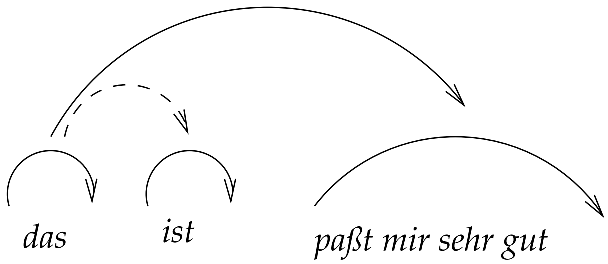


Abbildung 9: Chart für das Beispiel

Der Algorithmus muß so gestaltet sein, daß er beim Auffüllen der Chart adjazente Fragmente bevorzugt kombiniert, bevor Kombinationen vorgenommen werden, die Segmente überspringen. Daher wird zunächst die gestrichelte Kante (Abb. 9) eingetragen und eine (immer noch partielle) Repräsentation für *das ist* aufgebaut. Die Bewertung dieser Kante, die *das ist* überspannt, muß berücksichtigen:

1. Die Kante ist noch partiell (ein Argument des Verbs ist noch nicht gebunden).
2. Wenn die Prosodie hinter *ist* eine irreguläre Grenze erkannt hat, ist die resultierende Kante schlechter zu bewerten.

Dann wird durch Überspringen des Eintrags für *ist* eine Kante eingetragen, die *das* und *paßt mir gut* kombiniert. Bei ihrer Bewertung muß einfließen:

1. Eine andere Kante wird übersprungen (negativ); diese Kante ist kurz (ein Wort) und ist eventuell mit einer irregulären Grenze markiert (positiv).
2. Die resultierende Kante ist semantisch saturiert, da alle Argumentpositionen gebunden sind.

Diese Kante wird schließlich als Ausgabe ausgewählt, da sie die größte Abdeckung der Eingabe aufweist und semantisch vollständig saturiert ist.

4.3.4 Bewertungsstrategien

Um aus den berechneten Hypothesen eine Folge von Analysen auswählen zu können, die die gesamte Äußerung abdecken, müssen die Hypothesen bewertet werden.

Bei jeder Regelanwendung berechnet man dazu eine Bewertung für den neuen *chart*-Eintrag. Die Bewertung hängt u. a. von den folgenden Faktoren ab:

- Saturiertheit der semantischen Repräsentation, gegeben durch das Verhältnis der Zahl gebundener Argumente zu der Zahl der erwarteten Argumente,
- Abdeckung der Hypothese (Verhältnis der Länge des abgedeckten Zeitintervalls zur Gesamtlänge der Äußerung bzw. des zu verarbeitenden Segments),
- mit der angewendeten Regel verbundener Bonus oder Malus (um Regeln zu favorisieren oder zu unterdrücken),
- akustische Bewertung der Teilpfade, auf denen die partiellen Analysen basieren,
- Verletzung von sortalen Selektionsrestriktionen, usw.

Wie diese Faktoren gewichtet und zu einer Gesamtbewertung verrechnet werden, muß experimentell geklärt werden . Z. B. stellt sich die Frage, ob eine Analyse, die semantisch saturiert ist, ein Intervall der Eingabe aber nur teilweise abdeckt (d. h. es wurden Segmente ausgelassen), zu bevorzugen ist gegenüber einer Analyse, die größere Teile des Intervalls abdeckt, aber noch ungebundene Argumentpositionen enthält. Die Entwicklung der Bewertungsmaße erfolgt in Zusammenarbeit mit dem Arbeitspaket 2.2.3 *Holistische Suche* der Universität Erlangen-Nürnberg.

4.4 Anforderungen an andere Komponenten

Aus der dargestellten Funktionalität und der Einordnung in die Architektur des zukünftigen Verbmobil-Systems ergeben sich folgende Anforderungen an die Module, von denen die robusten semantischen Verfahren Eingaben benötigen:

Partielle Resultate: Die einzelnen Parser müssen in der Lage sein, einer Eingabeäußerung, die nicht vollständig analysierbar ist, eine Folge von partiellen Analysen zuzuordnen und diese an die robusten semantischen Verfahren weiterzugeben.

Resultate der verschiedenen Parser: Um aus den Resultaten mehrerer Analysekomponenten eine integrierte Analyse aufbauen zu können, ist für die robusten semantischen Verfahren der Zugriff auf die Ergebnisse aller

Parser erforderlich; das Arbeiten auf den fragmentarischen Ausgaben eines Parsers alleine dürfte eine geringere Verbesserung der Ausgabequaltität mit sich bringen.

Einhaltung des VIT-Formats: Alle Parser müssen Strukturen liefern, die der VIT-Definition genügen und in diesem Sinn wohlgeformt sind. Dies ist auch aus Sicht des Transfers und der Generierung eine essentielle Anforderung.

Prosodische Information: Abbrüche und Selbstkorrekturen sind z. T. anhand der prosodischen Markierung irregulärer Grenzen zu detektieren. Information über das Vorliegen solcher Markierungen muß den robusten semantischen Verfahren verfügbar sein. Bei der Detektion spontansprachlicher Phänomene wird mit dem Arbeitspaket 2.2.4 *Außergrammatische Phänomene* der Universität Erlangen-Nürnberg zusammengearbeitet.

Dialogaktvorhersagen: Anhand von Dialogaktvorhersagen lassen sich zahlreiche Entscheidungen bei der Integration fragmentarischer Analysen leichter treffen, etwa ob der Sprecher oder der Hörer als Füller für eine ungebundene Argumentrolle in Frage kommt.

4.5 Erste Erfahrungen aus der Sommerintegration 1997

Für die Sommerintegration 1997 wurde eine erste Version des Moduls *Semantics* erstellt und in das Gesamtsystem integriert. Neben den Resolutionsverfahren der Diskurs- und Dialogsemantik (Arbeitspaket 2.4.3) umfaßt es auch eine erste experimentelle Implementation der robusten semantischen Verfahren, allerdings nur mit rudimentärer Funktionalität. Das Ziel war dabei, die Interaktion mit der integrierten Verarbeitung und dem HPSG-Parser zu testen und einen Rahmen für die weitere Entwicklung bis zur nächsten Integration zu schaffen.

Es hat sich dabei gezeigt, daß die robusten semantischen Verfahren eine große Zahl partieller Analysen effizient verarbeiten müssen. Mit den Entwicklern des HPSG-Parsers wurde vereinbart, daß der Parser immer wenn er auf einem Pfad eine Phrase erkannt hat, den entsprechenden VIT über die integrierte Verarbeitung an das Modul *Semantics* schickt.

Dieses Protokoll führt zu einem großen Aufkommen an Analysen, die oft Teil vom Parser später gefundener und verschickter Ergebnisse sind. Wenn der Parser z. B. eine Nominalphrase findet, schickt er bereits den zugehörigen VIT ab. Wenn im weiteren Parsing diese Nominalphrase als Teil eines Satzes erkannt wird, wird ein VIT für den Satz verschickt. Dies ist zwar redundant, erlaubt aber eine inkrementelle Verarbeitung.

Als Hauptaufgaben für die weitere Entwicklung haben sich herausgestellt:

1. Die effiziente Handhabung einer großen Zahl partieller Analyseresultate und der durch Regelanwendung erzeugten Hypothesen,
2. die Formulierung von Regeln auf der Basis von konkreten Analyseergebnissen zu Corpusbeispielen.
3. die Entwicklung von Bewertungs- und Auswahlstrategien.

4.6 Vergleich mit anderen Ansätzen

Das in den Abschnitten 4.1-4.4 vorgestellte Modell zur Erhöhung der Robustheit des Verbmobil-Systems setzt wie die in Abschnitt 3.3 vorgestellten Arbeiten nach dem eigentlichen Parsing an. Es unterscheidet sich aber deutlich von den dort beschriebenen und auch von den anderen in 3.1 (vor dem Parsing) und 3.2 (während des Parsings) referierten Ansätzen. Seine wesentlichen Eigenschaften sind:

Verknüpfung von tiefen und flachen Analyseverfahren: Im Unterschied zu allen bekannten Ansätzen werden in dem vorgestellten Ansatz die Analyseresultate verschiedener Parser miteinander verknüpft. Die Ausgaben der Parser (HPSG-Parser, Chunk-Parser, statistischer Parser) sind untereinander vergleichbar und kombinierbar, da sie im VIT-Format geliefert werden.

Verbindung von Heuristiken und fundierter Semantik: Der Ansatz operiert auf linguistisch motivierten und interpretierbaren semantischen Objekten, den VITs. Diese semantischen Repräsentationen werden anhand von heuristischen Regeln und Verfahren miteinander kombiniert und die so aufgebauten Strukturen bewertet.

***Anytime*-Verhalten:** Aus der aufgebauten *chart* kann zu jeder Zeit eine optimale Hypothese extrahiert und ausgeben werden. Damit können die robusten semantischen Verfahren sich der zur Verfügung stehenden Zeit im Laufzeitsystem anpassen.

Inkrementalität: Inkrementell arbeitende Parser (wie z. B. der HPSG-Parser) können Zwischenresultate an die robusten semantischen Verfahren übergeben, die sofort in die *chart* eingetragen und weiterverarbeitet werden. Wenn die Analyse des inkrementellen Parsers fehlschlägt, können die robusten semantischen Verfahren bereits erste Resultate vorweisen.

Nutzung verschiedener Wissensquellen: Neben syntaktischem und semantischem Wissen wird auch prosodische und Diskursinformation in der Verarbeitung berücksichtigt.

5 Beispiele und Regeln

In diesem Abschnitt werden Beispiele[15] für die zu bearbeitenden Problemklassen im Detail diskutiert und exemplarisch zu ihrer Behandlung erforderliche Regeln in einer ersten rohen Fassung vorgestellt. Abschließend werden die für die einzelnen Probleme einschlägigen Wissensquellen und Regeltypen benannt.

5.1 Selbstkorrekturen und Abbrüche

5.1.1 Interjektionen als Korrekturindikatoren

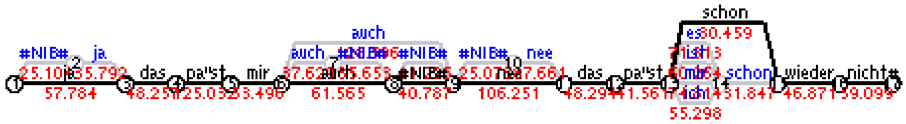
In einfachen Fällen nimmt der Sprecher eine Korrektur einer ganzen Proposition vor, indem er nach der Proposition eine korrekturmarkierende Interjektion und dann eine neue Proposition äußert, die seiner eigentlichen oder während des Sprechens geänderten Intention entspricht.

(7) Transliteration:[16]

```
das pa"st mir auch/- ach nee , das pa"st mir
schon wieder nicht .
```

(aus g191a.trl, KAE014)

(8) Wortgraph:



(9) Beste Kette:[17]

```
ja das pa"st mir auch #NIB# nee das pa"st
schon wieder nicht
```

[15] Die Beispiele sind dem Verbmobil-Corpus entnommen. Die Wörtergitter und die zugehörigen besten Ketten wurden auf den akustischen Originaldaten mit den im Forschungsprototyp 1.0 enthaltenen Spracherkennern erzeugt.

[16] In den Transliterationen markiert `/-` einen Abbruch und `+/ ... /+` eine Korrektur; wenn ein Wort abgebrochen wird, ist dies mit = markiert; `<Z>` steht für Zögern, `<P>` für eine Pause (Burger 1995; Burger 1997). Die anderen Annotationen sind für dieses Papier nicht relevant und können vom Leser ignoriert werden.

[17] `#NIB#` steht für *nichtinterpretierbarer Bereich*.

Wenn diese Äußerung durch Prosodie und Parser in drei Segmente aufgeteilt wird,[18] für den ersten Satz, die korrekturmarkierende Interjektion und den zweiten Satz, kann dieser Fall mit der folgenden einfachen Regel behandelt werden:

(10)
```
[type(V1,X),correction_marker(V2),type(V3,X)]
--->
[] & V3.
```

Diese Regel feuert, wenn in der Eingabe drei aufeinanderfolgende VITs auftreten, von denen der erste und der dritte einen identischen semantischen Typ haben (z. B. beide repräsentieren Propositionen oder Modifikatoren) und der zweite VIT eine Interjektion repräsentiert, die als Korrekturmarker interpretiert werden kann (Schmitz und Fischer 1995). Operationen werden beim Feuern dieses Regel nicht ausgeführt, ausgegeben wird der dritte VIT.

Zu den korrekturmarkierenden Interjektionen gehören *nein, nee, halt, stop, Moment* usw. Außer den korrekturmarkierenden Interjektionen müssen auch kurze phrasale Korrekturmarker wie *ich meine* berücksichtigt werden. Die Häsitationspartikel *äh* markiert zwar in einigen Fällen auch Korrekturen, überbrückt aber häufig auch nur eine Äußerungspause. Würde man sie auch als Korrekturmarker betrachten, so würde die Regel übergenerieren.

Die Regeln können nur als Heuristiken gewertet werden, deren Nützlichkeit nur durch Experimente mit Daten aus dem Corpus im laufenden Gesamtsystem (also mit Spracherkenner, Parser und robusten semantischen Verfahren) getestet werden kann. Durch Erkennerfehler kann es z. B. dazu kommen, daß eine Regel auf die Eingabe nicht anwendbar ist, obwohl das Phänomen, auf das sie abzielt, in der Äußerung vorliegt.

5.1.2 Korrektur von Modifikatoren

In der Verbmobil-Domäne kommen Zeit- und Datumsangaben sehr häufig vor. Dementsprechend oft treten auch Korrekturen bei Zeit- und Datumsausdrücken auf. Zur Behandlung des Beispiels *Sagen wir am ersten Juni äh am ersten Juli* (aus m027n, MPU019) ist eine Regel wie diese erforderlich:

(11)
```
[type(V1,prop),has_mod(V1,M1,ModType),
 type(V2,mod),has_mod(V2,M2,ModType)]
--->
[replace_mod(V1,M1,M2,V3)] & V3.
```

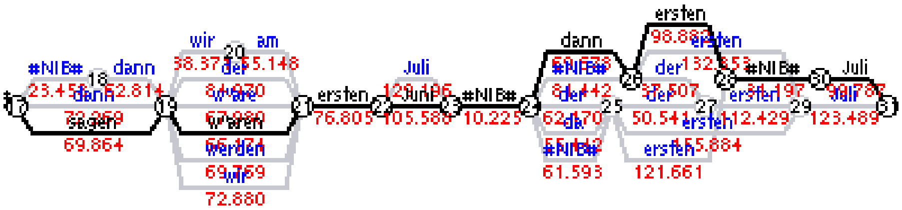
(12) Transliteration:

[18] Bei den Beispielen wird von einer Architektur ausgegangen, in der der Parser aufgrund prosodischer und syntaktischer Information die Eingabe segmentiert.

```
sagen wir +/am ersten Juni/+ <"ah> am ersten <P>
Juli .
```

(aus m027n, MPU019)

(13) Wortgraph:



(14) Beste Kette:

```
sagen w"aren ersten Juni #NIB# dann ersten #NIB# Juli
```

Angenommen, der Parser liefert für diese Äußerung die folgenden Segmente und analysiert *ersten Juni/Juli* als Modifikator:

(15) `[sagen wir ersten Juni][ersten Juli]`

Die Regel (11) besagt: Wenn in der Eingabe zwei VITs vorkommen, von denen der erste eine Proposition repräsentiert, die einen Modifikator enthält, und der zweite einen Modifikator vom gleichen Typ (z. B. temporal) und von gleicher Granularität (z. B. beides sind Datumsangaben wie im Beispiel) wie der im ersten VIT enthaltene, so ist der Modifikator aus dem ersten VIT zu entfernen und durch den zweiten Modifkator zu ersetzen. Das Ergebnis dieser Modifikation ist die Ausgabe dieser Regel. Die Regel kann noch dahingehend verfeinert werden, am Ende des ersten VITs nach einer irregulären prosodischen Grenze zu suchen.

5.1.3 Korrektur einer Fehlplanung

Bei Beispiel (16) hat der Sprecher zu einer Propositionalphrase angesetzt, dann aber offenbar bei der definiten Präposition das falsche Genus gewählt (*zum* statt *zur*), kann sie daher nicht zu Ende bringen und korrigiert sich:

(16) Transliteration:

```
dann schlag' ich vor , treffen wir uns doch +/zum/+
<A> zur Mittagszeit , so gegen ein Uhr ?
```

(aus m095d, DAN012)

Dieses Beispiel zeigt die Vorteile der Kopplung der robusten semantischen Verfahren mit einem inkrementellen Parser (wie dem HPSG-Parser im Verbmobil 2-System). Angenommen, der Parser findet zunächst die Phrasen *dann schlage ich vor*, *treffen wir uns doch* und *zur Mittagszeit* sowie die einzelne Präposition *zum*. Außerdem wird er später die Phrase *dann schlage ich vor, treffen wir uns doch* finden.

(17)
```
[type(V1,mod),type(V2,prop)]
--->
[apply(V1,V2,V3)] & V3.
```

Die Regel (17), die Modifikatoren an Propositionen anbindet, kann unter überspringen des *zum* das *zur Mittagszeit* an das *treffen wir uns doch* anbinden. Das Resultat wird dann als propositionales Argument an *dann schlag ich vor* angebunden.

Allerdings ergibt sich bei diesem Beispiel eine Ambiguität, da der Modifikator *zur Mittagszeit* natürlich auch an *dann schlage ich vor, treffen wir uns doch* angebunden werden kann.

5.1.4 Prononima als abgebrochenes Satzpräfix

Sehr häufig kommt es vor, das zu einem Satz angesetzt wird, bei dem ein Pronomen in topikalisierter Position steht, aber nicht mehr als dieses Pronomen geäußert wird, sondern ein neuer vollständiger Satz angeschlossen wird:

(18) Transliteration:

```
+/das/+ also das k"onnen wir dann<Z> festhalten ,
denke ich .
```

(aus g072a, TIS011)

(19)
```
[type(V1,anaphor),type(V2,prop)] --->
[] & V2.
```

Durch die Regel (19), die allein vorkommende anaphorische Ausdrücke unmittelbar vor Propositionen tilgt, lassen sich diese Fälle behandeln.

Dieses Beispiel verdeutlicht auch, daß in den Daten nicht immer eindeutig zwischen Selbstkorrekturen und Abbrüchen unterschieden werden kann. Nach der Transliteration handelt es sich um eine Selbstkorrektur; nach den in Abschnitt 2.1.2 genannten Kriterien wäre dieses Beispiel eher als Abbruch zu klassifizieren.

5.1.5 Eine abgebrochene Frage

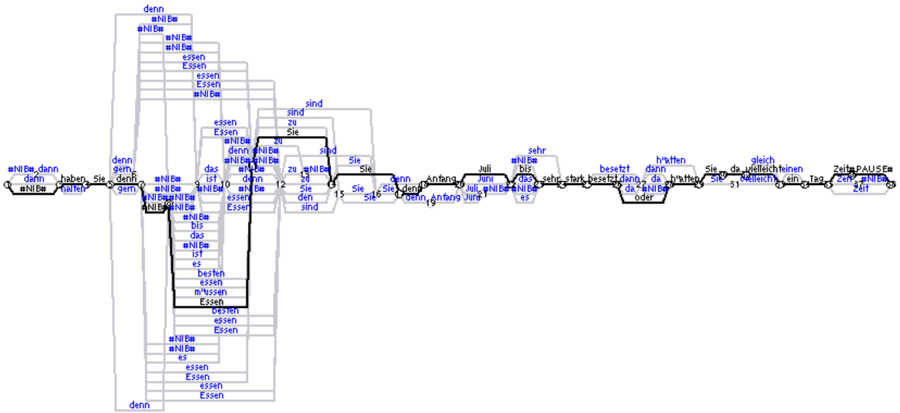
In diesem Beispiel setzt der Sprecher zu einer Frage an, bricht sie nach drei Wörtern ab und fährt mit einer neuen Frage fort.

(20) Transliteration:

```
<A> <"ahm> +/haben Sie denn<Z>/+ sind Sie denn Anfang
Juli<Z> <P> sehr stark besetzt , oder h"atten Sie da
vielleicht einen Tag Zeit <A> ?
```

(aus g094a, ANS010)

(21) Wortgraph:



(22) Beste Kette:

```
#NIB# haben Sie denn #NIB# essen Sie Sie denn Anfang
Juli bis sehr stark besetzt oder h"atten Sie da
vielleicht ein Tag Zeit #PAUSE#
```

Bei diesem Beispiel ist damit zu rechnen, daß der Parser den abgebrochenen Teil der Äußerung nur als einzelne Wörter analysieren wird, während der neuangesetzte Satz eine vollständige Analyse erhält:

(23)
```
[haben][Sie][denn][sind Sie denn Anfang Juli sehr
stark besetzt oder hätten Sie vielleicht einen
Tag Zeit]
```

Zunächst wird *haben* mit *Sie* kombiniert, auf das (immer noch partielle) Ergebnis wird der Modifikator *denn* angewendet. Dann kann dieses abgebrochene Präfix durch diese Regel ausgefiltert werden:

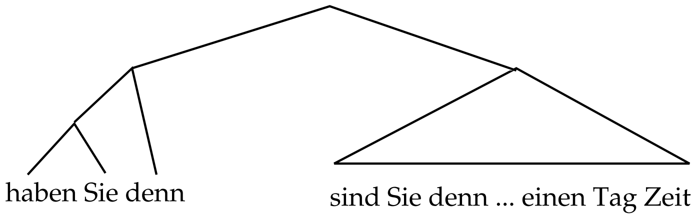


Abbildung 10: Integration des Beispiels für einen Abbruch

(24)
```
[[unsaturated(V1),beginning_of_segment(V1)],
 type(V2,prop)] --->
[] & V2
```

Die Regel feuert, wenn am Anfang eines Segments ein VIT gefunden wird, der nicht saturiert ist (d. h. es fehlen noch obligatorische Argumente) auf den unmittelbar eine vollständige Proposition folgt. Die Regel kann noch erweitert werden, um prosodische Information zu berücksichtigen, wenn die prosodische Analyse Abbruchstellen mit ausreichender Sicherheit erkennen kann.

5.1.6 Korrektur eines Verbs

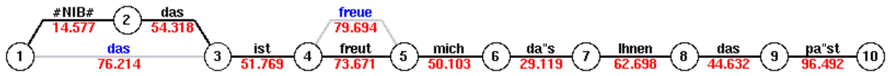
In dieser Äußerung setzt der Sprecher den Satz anders fort, als er ihn begonnen hat, indem er das bereits (teilweise) geäußerte Verb *ist* durch das Verb *freut* korrigiert:

(25) Transliteration:

```
das +/is<Z>=/+ freut mich , da"s Ihnen das pa"st <A>
<#Klicken> .
```

(aus g097a, ANS003)

(26) Wortgraph:



(27) Beste Kette:

```
#NIB# das ist freue mich da"s Ihnen das pa"st
```

Angenommen, der Parser liefert für diese Äußerung die folgenden drei Segmente:

(28) `[das][ist][freut mich da"s Ihnen das pa"st]`

d. h. das abgebrochene +/is<Z>=/+ wird vom Spracherkenner als *ist* erkannt.

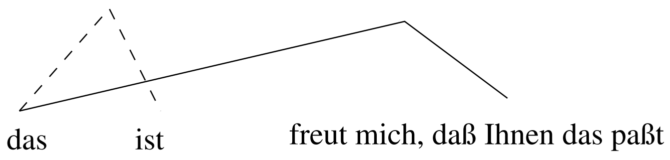


Abbildung 11: Integration des Beispiels für eine Verbkorrektur

Bei diesem Beispiel sollte das *ist* unterdrückt werden und *das* mit dem Rest der Äußerung, *freut mich, daß Ihnen das paßt*, kombiniert werden.

Für die Behandlung dieses Beispiels sind zwei Varianten denkbar. Einerseits muß bei der Abarbeitung der Regeln die Möglichkeit vorgesehen werden, einzelne Eingabesegmente zu überspringen, wenn ansonsten keine Regelanwendung möglich ist. Dann kann die VP als Funktor auf das Pronomen als Argument angewendet werden.[19]

Die andere Möglichkeit bezieht prosodische Information ein. Laut Transliteration hat der Sprecher nach dem Wortfragment gezögert: +/is<Z>=/+. Wenn die Prosodie an dieser Stelle eine irreguläre Grenze erkennt, kann dieses Segment aufgrund dieser Information herausgefiltert werden:

(29)
```
[type(V1,_),irreg_boundary_at_end(V2)] --->
[] & V1.
```

5.2 Erkennerfehler

5.2.1 Ein komplexes Beispiel für Erkennerfehler

Wenn in der Ausgabe des Spracherkenners nur einzelne Wörter falsch sind, kann dies bereits dazu führen, daß die Eingabe nicht als ganzes parsbar ist. In diesem Beispiel wurden zwei Wörter durch andere ersetzt; eine der beiden Ersetzungen (*dort* → *doch*) wäre für die Analysierbarkeit unschädlich, da ein Adverb durch ein Adverb ersetzt wird, die andere (*werde* → *wir*) läßt das Parsing wegen eines Kategorienwechsels aber scheitern.

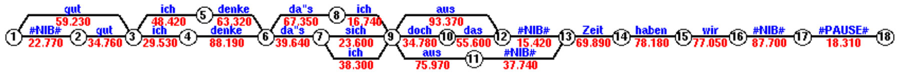
(30) Transliteration:

[19]Unter der Annahme, daß *das* hier als Argument und nicht als Expletiv analysiert wird.

```
gut , ich denke , da"s ich dort Zeit haben werde .
```

(aus n016k, OC1014)

(31) Wortgraph:



(32) Beste Kette:

```
#NIB# gut ich denke da"s sich doch das #NIB# Zeit
haben wir #NIB# #PAUSE#
```

Angenommen der Parser liefert Analysen für folgende Segmente:[20]

(33) `[gut][ich denke][da"s][ich][doch][das][Zeit][haben][wir]`

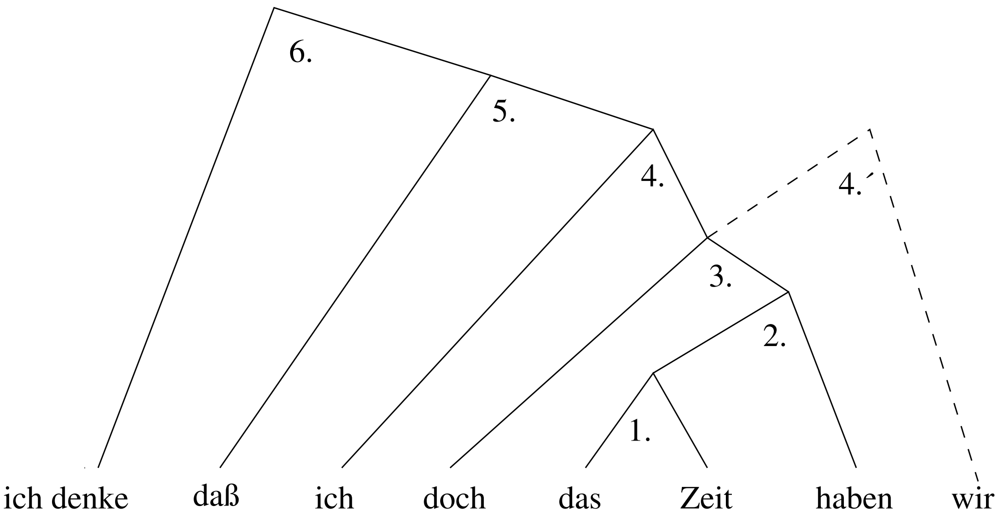


Abbildung 12: Integration des Beispiels

In Abb. 12 ist die Reihenfolge, in der die Segmente idealerweise kombiniert werden, dargestellt. An den Knoten werden folgende Heuristiken[21] angewendet:

[20] Dies ist eine vereinfachende Annahme, da *sich* vom Erkenner besser bewertet wird als *ich*.

[21] Diese Regeln werden nur grob skizziert; teilweise wird man sie durch abstraktere Regeln ersetzen können, teilweise wird man sie stärker einschränken müssen.

1. Kombiniere einen Determiner oder Quantoren mit einem Prädikat: Aus `[das]` und `[Zeit]` wird eine definite NP. Da diese Kombination auf semantischer Ebene passiert, wird von der verletzten Kongruenz abstrahiert. Dies wird erreicht durch die oben bereits eingeführte Regel, die einen Funktor, dem noch ein Argument fehlt, mit einem potentiellen Argument kombiniert:

   ```
   [unbound_arg(V1,Type),type(V2,Type)] --->
   [apply(V1,V2,V3)] & V3.
   ```

   Diese Regel wird auch in den folgenden Schritten angewendet. Die Verletzung der Kongruenz könnte bei der Bewertung der Hypothese als negativ berücksichtigt werden.

2. Kombiniere ein verbales Prädikat, dem noch Argumente fehlen, mit einem potentiellen Argument (um zu prüfen, ob eine solche Kombination sinnvoll ist, können Sortentests und Statistiken über Kombinationswahrscheinlichkeiten einbezogen werden).

3. Wende einen Modifikator auf ein potentielles Argument an (Adverb auf VP, Adjektiv auf N, wieder sind Tests möglich).

4. s. 2. (4.' kann durch eine in die Suchstrategie integrierte Meta-Heuristik „Arbeite bevorzugt von links nach rechts“ ausgeschlossen werden).

5. Kombiniere einen Komplementierer, eine subordinierende Konjunktion oder ein Diskursrelationsadverb mit einem Satzkomplement.

6. Kombiniere einen Satz, der ein Einstellungsverb ohne ein propositionales Komplement enthält, mit einem durch Komplementierer eingeleiteten Satz. *Oder einfacher wieder:* Kombiniere ein verbales Prädikat, dem noch Argumente fehlen, mit einem potentiellen Argument.

Als Resultat erhält man einen VIT, der sich paraphrasieren läßt als *Ich denke, daß ich doch die Zeit habe.*

5.2.2 Erkennerfehler und Diskursinformation

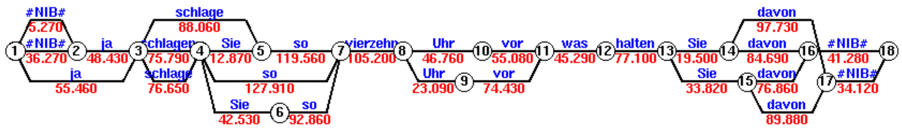
In diesem Beispiel läßt der Erkenner das Subjekt des Satzes, *ich*, aus. Hier kann eine Heuristik Abhilfe schaffen, die in einem solchen Fall einen Füller der Argumentrolle einträgt. Was als Füller in Frage kommt, kann vom vorhergesagten Dialogakt abhängen; bei einem Vorschlag ist der Sprecher eine plausible Alternative.

(34) Transliteration:

```
<A> ja , ich schlage so<Z> vierzehn Uhr vor .
was halten Sie davon ? <#Klicken>
```

(aus n022k, TS2010)

(35) Wortgraph:



(36) Beste Kette:

```
 #NIB# ja schlagen so vierzehn Uhr vor was halten Sie
davon #NIB#
```

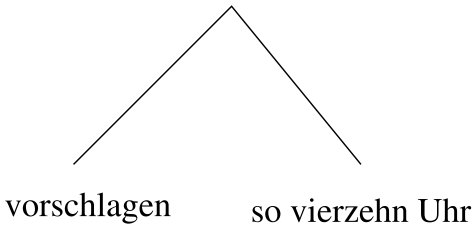


Abbildung 13: Integration des zweiten Beispiels

Angenommen, der Parser liefert folgende Segmente:

(37)
```
[ja][schlagen][so vierzehn Uhr][vor]
[was halten Sie davon]
```

Es sei vereinfachend angenommen, daß das abtrennbare Präfix *vor* in einem Vorverarbeitungsschritt als solches identifiziert und das Verb als *vorschlagen* interpretiert wird.

- *ja* kann zu Beginn eines Turns nur eine Interjektion sein (aber kein Adverb) und wird daher durchgereicht.
- *so vierzehn Uhr* kann als direktes Objekt von *vorschlagen* ermittelt werden.

```
[unbound_arg(V1,Type),type(V2,Type)] --->
[apply(V1,V2,V3)] & V3.
```

- Das Subjekt von *vorschlagen* fehlt; der Sprecher kann als Subjekt eingetragen werden, wenn die Dialogaktvorhersage einen Vorschlag (oder einen ähnlichen Dialogakt) voraussagt.

```
[type(V1,verbal),missing_arg(V1,Arg,human),
 dialogue_act_prediction(suggestion)]
--->
[fill_arg_slot(V1,Arg,speaker,V2)] & V2.
```

- *was halten Sie davon* wird durchgereicht.

Das Ergebnis sind VITs für *Ja. Ich schlage so vierzehn Uhr vor. Was halten Sie davon?*

5.2.3 Erkennerfehler und Typanpassung

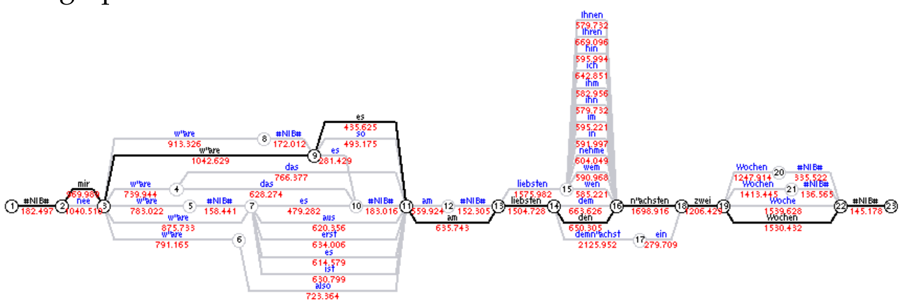
Bei diesem Beispiel tritt das Problem auf, daß die korrekte Wortfolge *in den nächsten zwei Wochen* nicht erkannt wird, sondern *in* und *den* werden in überlappenden Intervallen erkannt.

(38) Transliteration:

```
mir waere es am liebsten in den naechsten zwei Wochen
```

(aus den Moko-Testdaten)

(39) Wortgraph:



(40) Beste Kette:

```
#NIB# mir waere es am liebsten den naechsten
zwei Wochen #NIB#
```

Der Parser liefert voraussichtlich die folgenden Segmente zu dieser Eingabe:

(41) `[mir wäre es am liebsten][den nächsten zwei Wochen]`

Temporale Nominalphrasen wie *die nächste Woche*, *den Montag* usw. kommen im Verbmobil-Corpus häufig als Modifikatoren vor, z. B. in *Wir treffen uns die nächste Woche*. Daher und zur Behandlung des genannten Beispiel ist es sinnvoll, eine Regel zu formulieren, die temporale Nominalphrasen als Modifikatoren uminterpretiert, so daß sie auf potentielle Modifikanden angewendet werden können. Vereinfachend wird dabei angenommen, daß die rekonstruierte Modifikationsbeziehung eine temporale Lokalisierung des Modifikanden ist.

(42)
```
[temporal_np(V1)] --->
[typeraise_to_mod(V1,V2)] & V2.
```

Im genannten Beispiel würde diese Regel auf *den nächsten zwei Wochen* angewendet. Der resultierende Modifikator kann dann durch die bereits eingeführte Regel (17) zur Modifikatorenapplikation angewendet werden auf *mir wäre es am liebsten*.

5.3 Segmentierungsprobleme und unzureichende Abdeckung der Grammatik

Dieses Beispiel (das nicht aus dem Verbmobil-Corpus stammt) wird von Rosé (1996) diskutiert (vgl. 3.3.1). Es kann als Beispiel sowohl für Segmentierungsprobleme als auch für eine unzureichende Abdeckung der Grammatik dienen. Der Eingabesatz lautet *Thursday, I am out.* Wenn die Grammatik topikalisierte Modifikatoren nicht zuläßt, liefert der Parser folgende Segmente:

(43) `[Thursday][I am out]`

Ein alleinstehender Wochentag wird von der Grammatik wahrscheinlich als Modifikator analysiert; wenn er als Nomen analysiert wird, kann er durch *type raising* als Modifikator uminterpretiert werden. Eine Korrekturregel kann dann diesen Modifikator mit einer benachbarten Proposition kombinieren.

(44)
```
[type(V1,mod),type(V2,prop)]
--->
[apply(V1,V2,V3)] & V3.
```

Die Anwendbarkeit dieser Regel muß aber so eingeschränkt werden, daß sie im Falle einer Linksversetzung nicht *Am Montag ... das paßt mir gut* auf *Das paßt mir gut am Montag* abbildet. Eine Möglichkeit dazu besteht darin, die Kombination von Fragmenten auszuschließen, wenn zwischen ihnen mit hoher Wahrscheinlichkeit eine prosodische Grenze erkannt wurde.

5.4 Wissensquellen und Regeln

Die Wissensquellen und Regeltypen, die zur Behandlung der Beispiele in den vorigen Abschnitten verwendet wurden, lassen sich den verschiedenen Problemklassen zuordnen.

5.4.1 Verwendete Wissensquellen

Für die Behandlung der in den vorigen Abschnitten beschriebenen Probleme sind unterschiedliche Wissensquellen relevant (s. Abb. 14). Für alle Problemgruppen wird als Informationsquelle syntaktisches und semantisches Wissen herangezogen.

Problem	Syntax/Semantik: Valenz	Syntax/Semantik: sortale Restriktionen	Prosodie	Dialogakte
Abbrüche	+	−	+	−
Selbstkorrekturen	+	+	+	−
Erkennerfehler	+	+	−	+
Segmentierung	+	+	+	+
Grammatikabdeckung	+	+	−	−

Abbildung 14: Wissensquellen zur Behandlung der einzelnen Probleme

Da die zentrale Aufgabe der Umgang mit partiellen Analysen und das Kombinieren solcher Fragmente ist, bildet Valenzinformation den Kern des notwendigen Wissens für alle Problemklassen, da sie die mögliche Kombinatorik der Fragmente beschreibt. Gewonnen wird die Information aus den syntaktischen und semantischen Lexika der einzelsprachlichen Grammatiken. Anhand der Valenz läßt sich die Unvollständigkeit einer abgebrochenen Äußerung oder eines korrigierten Äußerungsanteils erkennen.

Bei Fragmentierung durch Erkennerfehler, einer zu feinkörnigen Segmentierung oder Problemen der Grammatikabdeckung werden Fragmente anhand ihrer Valenz kombiniert. Zur Beschränkung der Kombinatorik dienen sortale Selektionsrestriktionen.

Information über (irreguläre) prosodische Grenzen ermöglicht die Erkennung spontansprachlicher Phänomene und die Ermittlung des Umfangs eines abgebrochenen oder korrigierten Äußerungsanteils. Bei Segmentierungsproblemen ist sie weniger hilfreich, zumal sie hier u. U. die Ursache des Problems war. Auch bei grammatischen Äußerungen außerhalb der Abdeckung der Grammatik kann Prosodie nicht helfen.

Die Verfügbarkeit vorgesagter Dialogakte unterstützt die Behandlung von Erkennerfehlern und Segmentierungsproblemen, da anhand des Dialogakts

Hypothesen über Satzmodus usw. aufgestellt werden können.

5.4.2 Regelklassen

Die in den vorigen Abschnitten verwendeten Regeln lassen sich aufteilen in zwei Gruppen:

1. Tilgungsregeln und
2. Kombinationsregeln.

Tilgungsregeln sind dadurch gekennzeichnet, daß sie komplette Eingabesegmente aus der weiteren Verarbeitung entfernen. Dies ist nötig bei abgebrochenen Fragmenten und den korrigierten Teilen einer Selbstkorrektur.

(45)
```
[type(V1,X),correction_marker(V2),type(V3,X)]
--->
[] & V3.
```

Diese in Abschnitt 5.1 eingeführte Regel bildet drei Eingabesegmente auf eines ab, indem sie zwei der Segmente löscht.

Kombinationsregeln fügen mehrere partielle Analysen zu einer Analyse mit größerer Abdeckung zusammen. Ein Beispiel ist die Regel, die ein noch ungebundenes Argument eines Funktors abbindet (s. a. (6), S. 30).

(46)
```
[unbound_arg(V1,Type),type(V2,Type)] --->
[apply(V1,V2,V3)] & V3.
```

Diese Regel konsumiert zwei partielle Analysen und liefert eine dritte, die durch Applikation der einen (Funktor) auf die andere (Argument) entstanden ist.

Tilgungsregeln finden Anwendung bei spontansprachlichen Phänomenen, Kombinationsregeln werden bei allen Problemklassen angewendet.

5.5 Schwierigkeitsgrad

Die diskutierten Problemklassen, spontansprachliche Phänomene, Erkennerfehler, Segmentierungsproblem und unzureichende Abdeckung der Grammatik, unterscheiden sich in ihrem Schwierigkeitsgrad sowie in den Aussichten, durch die robusten semantischen Verfahren erfolgreich behandelt werden zu können.

Am besten sind die Erfolgsaussichten bei spontansprachlichen Phänomenen und Erkennerfehlern, die auch die häufigsten Probleme darstellen. Im Rahmen des Arbeitspaketes werden zunächst diese Problemklassen bearbeitet, da die spontansprachlichen Phänomene unabhängig von den Parsern und

Grammatiken auftreten und Erkennerfehler jetzt und in Zukunft ein Problem sind.

Da für die Erkennung spontansprachlicher Phänomene die Detektion irregulärer prosodischer Grenzen eine wichtige Informationsquelle ist, können diese Phänomene aber erst dann erfolgreich bearbeitet werden, wenn die prosodische Analyse Information über diese Art von Grenzen liefern kann.[22]

Zunächst wird im Verbmobil 2-System nur ein Parser, der tiefe HPSG-Parser, eingesetzt. Dieser ist in der Lage, auf nicht vollständig analysierbaren Eingaben partielle Analysen zu liefern, so daß auf der Basis dieses Parser die robusten semantischen Verfahren in erster Version entwickelt werden können. Mit der Integration der anderen Parser in das System wird das Datenaufkommen und das Volumen der Eingaben für die robusten semantischen Verfahren steigen; diese Erweiterung stellt somit einen Test für die Skalierbarkeit des Konzepts dar. Gleichzeitig steigen mit der Verfügbarkeit von Eingaben aus verschiedenen Analyseverfahren aber auch die Erfolgschancen für die robusten semantischen Verfahren.

Segmentierungsprobleme und unzureichende Abdeckung der Grammatik werden in der Bearbeitung vorläufig zurückgestellt. Die Grammatiken sind noch in der Entwicklung; wenn unabhängig davon formulierte allgemeine Regeln der robusten semantischen Verfahren hier Abdeckungslücken ausgleichen, ist dies ein willkommener Nebeneffekt. Um Probleme der Segmentierung zu bearbeiten, müssen zunächst Erfahrungen mit der prosodischen Analyse und den Parsern sowie den ihnen zugrundeliegenden Grammatiken gesammelt werden.

Allgemein muß angemerkt werden, daß es sich bei den robusten semantischen Verfahren um ein experimentelles Vorhaben handelt. Ihr Nutzen und ihre Anwendbarkeit muß laufend durch Experimente auf Corpus-Daten überprüft und das Vorgehen ggf. modifiziert werden.

[22] Im Rahmen des Arbeitspakets 2.2.4 *Außergrammtische Phänomene* wird an der Universität Erlangen-Nürnberg untersucht, inwieweit sich spontansprachliche Phänomene anhand prosodischer Information und unter Verwendung von stochastischen Modellen erkennen lassen.

6 Zusammenfassung

Das Ziel der robusten semantischen Verfahren ist es, die Robustheit des Verbmobil-Systems gegenüber spontansprachlichen Phänomenen und Erkennerfehlern zu erhöhen und Segmentierungsprobleme und Lücken in der Grammatikabdeckung abzumildern. Die Vorgehensweise besteht darin, bei einer durch diese Phänomene verursachten fragmentierten Analyse die partiellen Analyseresultate zu größeren Einheiten zusammenzufügen.

Eine wesentliche Eigenschaft, die dies ermöglicht, ist die Verknüpfung der Resultate unterschiedlicher tiefer und flacher Analyseverfahren in einer *chart*-Datenstruktur auf der Ebene des gemeinsamen VIT-Formats. Die Nutzung der Ergebnisse der verschiedenen Parsingverfahren bewirkt, daß deren jeweilige Stärken gebündelt werden. Eine solche Kombination der Resultate verschiedener Analyseverfahren ist bisher noch nicht vorgeschlagen und erprobt worden.

Informationen aus unterschiedlichen Wissensquellen (Syntax, Semantik, Prosodie, Diskurs- und Kontextinformation) werden in dem vorgestellten Modell zur Verarbeitung fragmentierter Analysen verwendet. Es kombiniert linguistisch fundierte Analyseverfahren und Repräsentationformate für Äußerungsbedeutungen mit heuristischen Regeln und eignet sich auch für ein inkrementelles Verarbeitungsmodell.

Ein weiteres wichtiges Merkmal ist, daß aus der aufgebauten *chart*-Datenstruktur jederzeit eine beste Hypothese extrahiert werden kann (*anytime*-Verhalten). Damit können sich die robusten semantischen Verfahren den verfügbaren Ressourcen im Laufzeitsystem optimal anpassen.

Literatur

Antoine, J.-Y. (1996). Parsing spoken language without syntax. In *Proc. of the 16th COLING*, Copenhagen, Denmark, pp. 47–52.

Bear, J., J. Dowding, und E. Shriberg (1992). Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog. In *Proc. of the 30 th ACL*, Newark, DE, pp. 56–63.

Block, H. U. (1997). The language components in Verbmobil. Vgl. IEEE Signal Processing Society (1997), pp. 79–82.

Block, H.-U. und S. Schachtl (1995). What a grammar of spoken dialogues has to deal with. In G. Heyer und H. Haugeneder (Hrsg.), *Language Engineering*, Advanced Studies in Computer Science, pp. 101–126. Braunschweig, Wiesbaden: Vieweg.

Boddy, M. (1991). Anytime problem solving using dynamic programming. In *Proc. of AAAI-91*, Anaheim, CA, pp. 738–743.

Boros, M., W. Eckert, F. Gallwitz, G. Görz, G. Hanrieder, und H. Niemann (1996). Towards understanding spontaneous speech: Word accuracy vs. concept accuracy. In *Proc. of the 1996 International Conference on Spoken Language Processing (ICSLP 94)*, Philadelphia, USA, pp. 1005–1008.

Bos, J., B. Gambäck, C. Lieske, Y. Mori, M. Pinkal, und K. Worm (1996). Compositional semantics in Verbmobil. In *Proc. of the 16th COLING*, Copenhagen, Denmark, pp. 131–136.

Bos, J., M. Schiehlen, und M. Egg (1996, August). Definition of the abstract semantic classes for the Verbmobil Forschungsprototyp 1.0. Verbmobil-Report 165, Univ. des Saarlandes, Univ. Stuttgart, IBM Deutschland.

Bub, T., W. Wahlster, und A. Waibel (1997). Verbmobil: The combination of deep and shallow processing for spontaneous speech translation. Vgl. IEEE Signal Processing Society (1997), pp. 71–74.

Burgard, C., R. Karger, und W. Wahlster (Hrsg.) (1996). Wissenschaftliche Ziele und Netzpläne für Verbmobil Phase 2. Verbmobil Technisches Dokument 44, DFKI GmbH, Saarbrücken.

Burger, S. (1995, Oktober). Transliterationslexikon. Verbmobil Technisches Dokument 36, Ludwig-Maximilians-Universität München, Institut für Phonetik und sprachliche Kommunikation, München.

Burger, S. (1997, April). Transliteration spontansprachlicher Daten. Verbmobil Technisches Dokument 56, Ludwig-Maximilians-Universität München, Institut für Phonetik und sprachliche Kommunikation, München.

Carter, D., R. Becket, M. Rayner, R. Eklund, C. MacDermid, M. Wirén, S. Kirchmeier-Andersen, und C. Philp (1997). Translation methodology in the Spoken Language Translator: An evaluation. In *Proc. of the 35 th ACL*, Madrid, Spain, pp. to appear. Workshop on Spoken Language Translation.

Cormen, T. H., C. E. Leiserson, und R. L. Rivest (1990). *Introduction to Algorithms*, Chapter 16, Dynamic Programming, pp. 301–328. Cambridge, MA, London, England. MIT Press.

Dean, T. und M. Boddy (1988). An analysis of time-dependent planning. In *Proc. of AAAI-88*, St. Paul, MN, pp. 49–54.

Dorna, M. und M. C. Emele (1996). Efficient implementation of a semantic-based transfer approach. In *Proc. of the 12 th ECAI*, Budapest, Hungary.

Eckert, W. und H. Niemann (1994). Semantic analysis in a robust spoken dialog system. In *Proc. of the 1994 International Conference on Spoken Language Processing (ICSLP 94)*, Yokohama, Japan, pp. 107–110.

Ehrlich, U. und G. Hanrieder (1996). Robust speech parsing. In J. Carroll (Hrsg.), *Workshop on Robust Parsing*, Prague, pp. 26–34. Eighth European Summer School in Logic, Language and Information.

Finkler, W. (1997). *Automatische Selbstkorrektur bei der inkrementellen Generierung gesprochener Sprache unter Realzeitbedingungen. Ein empirisch-simulativer Ansatz unter Verwendung eines Begründungsverwaltungssystems*. Dissertation, Universität des Saarlandes, Technische Fakultät.

Frederking, R. und S. Nirenburg (1994). Three heads are better than one. In *Proc. of the 4 th Conference on Applied Natural Language Processing*, Stuttgart, Germany, pp. 95–100.

Gazdar, G. und C. Mellish (1989). *Natural Language Processing in PROLOG*. Wokingham, England: Addison-Wesley Publishing Company.

Hanrieder, G. und G. Görz (1995). Robust parsing of spoken dialogue using contextual knowledge and recognition probabilities. In *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, Vigsø, Denmark.

Hayes, P. J., A. G. Hauptmann, J. G. Carbonell, und M. Tomita (1986). Parsing spoken language: A semantic caseframe approach. In *Proc. of the 11 th COLING*, Bonn, FRG, pp. 587–592.

Heeman, P. und J. Allen (1994). Detecting and correcting speech repairs. In *Proc. of the 32 th ACL*, Las Cruces, NM, pp. 295–302.

Heeman, P. A. und J. Allen (1997). Intonational boundaries, speech repairs and discourse markers: Modeling spoken dialog. In *Proc. of the 35 th ACL*, Madrid, Spain, pp. to appear.

Hindle, D. (1983). Deterministic parsing of syntactic non-fluencies. In *Proc. of the 21 th ACL*, Cambridge, MA, pp. 123–128.

Hipp, D. R. (1993). *Design and Development of Spoken Natural-Language Dialog Parsing Systems*. Ph. D. thesis, Department of Computer Science, Duke University, Durham, NC.

Höhne-Leska, C. (1975). *Statistische Untersuchungen zur Syntax gesprochener und geschriebener deutscher Gegenwartssprache*. Band 59, Heft 1 der Abhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig, Philologisch-historische Klasse. Berlin: Akademie-Verlag.

IEEE Signal Processing Society (1997). *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, München, Germany. IEEE Signal Processing Society.

Kompe, R., A. Kießling, H. Niemann, E. Nöth, A. Batliner, S. Schachtl, T. Ruland, und H. U. Block (1997). Improving Parsing of Spontaneous Speech with the Help of Prosodic Boundaries. Vgl. IEEE Signal Processing Society (1997), pp. 75–78.

Langer, H. (1990). Syntactic normalization of spontaneous speech. In *Proc. of the 13 th COLING*, Helsinki, Finnland, pp. 180–183.

Lavie, A. (1996). *GLR*: A Robust Grammar-Focused Parser for Spontaneously Spoken Language*. Ph. D. thesis, Carnegie Mellon University, School of Computer Science.

Lavie, A., D. Gates, M. Gavaldà, L. Mayfield, A. Waibel, und L. Levin (1996). Multi-lingual translation of spontaneously spoken language in a limited domain. In *Proc. of the 16th COLING*, Copenhagen, Denmark, pp. 442–447.

Lehning, M. (1996, August). Evaluierung von signalnahen Spracherkennungssystemen für die deutsche Spontansprache. Verbmobil-Report 161, Technische Universität Braunschweig, Institut für Nachrichtentechnik.

Levelt, W. J. M. (1983, Juli). Monitoring and self-repair in speech. *Cognition 14*(1), 41–104.

Levelt, W. J. M. (1989). *Speaking. From Intention to Articulation*. Cambride, MA, und London, England: MIT Press.

Mayfield, L., M. Gavaldà, W. Ward, und A. Waibel (1995). Concept based speech translation. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, Detroit, MI.

Nakatani, C. und J. Hirschberg (1993). A speech-first model for repair detection and correction. In *Proc. of the 31 th ACL*, Columbus, OH, pp. 46–53.

Nederhoff, M.-J., G. Bouma, R. Koeling, und G. van Noord (1997). Grammatical analysis in the OVIS spoken-dialogue system. In *Proc. of the 35 th ACL*, Madrid, Spain, pp. to appear. Workshop on Spoken Dialogue Systems.

Nöth, E. und B. Plannerer (1993, Dezember). Schnittstellendefinition für den Worthypothesengraphen. Verbmobil-Memo 2, Universität Erlangen-Nürnberg und Technische Universität München.

Oerder, M. und H. Ney (1993). Word graphs: An efficient interface between continuous-speech recognition and language understanding. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Wo??? IEEE Signal Processing Society.

Quantz, J. J. und J. Alexandersson (1996). Word lattice parsing based on heterogeneous information sources. In *Proc. of the 12 th ECAI*, Budapest, Hungary.

Rayner, M. und D. Carter (1997). Hybrid language processing in the spoken language translator. Vgl. IEEE Signal Processing Society (1997), pp. 107–110.

Reinecke, J. (1996, November). Evaluierung der signalnahen Spracherkennung im Verbundprojekt Verbmobil (Herbst 1996). Verbmobil-Memo 113, Technische Universität Braunschweig, Institut für Nachrichtentechnik.

Rosé, C. P. (1996). A genetic programming approach to robust interactive discourse interpretation. In *Procceedings of the AAAI-96 Workshop on on Detecting, Repairing, and Preventing Human-Machine Miscommunication*, Portland, Oregon.

Rosé, C. P. (1997). *Robust Interactive Dialogue Interpretation*. Ph. D. thesis, Carnegie Mellon University, Pittsburgh, PA. welche School?

Rosé, C. P. und A. Lavie (1997). An efficient distribution of labor in a two stage robust interpretation process. In *Proc. of the Second Conference On Empirical Methods in Natural Language Processing (EMNLP-2)*.

Rosé, C. P. und A. H. Waibel (1996). Recovering from parser failures: A hybrid statistical and symbolic approach. In J. L. Klavans und P. Resnik (Hrsg.), *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, Language, Speech, and Communication, Chapter 8, pp. 157–179. Cambridge, MA, London, England: The MIT Press.

Samuelsson, C. (1994, Februar). *Fast Natural Language Parsing Using Explanation-Based Learning*. Ph. D. thesis, The Royal Institute of Technology (KTH), Department of Computer and System Sciences, Stockholm, Sweden.

Schmitz, B. und K. Fischer (1995, Juni). Pragmatisches Beschreibungsinventar für Diskurspartikel und Routineformeln anhand der Demonstratorwortliste. Verbmobil-Memo 75, Universität Bielefeld und Technische Universität Berlin.

Stallard, D. und R. Bobrow (1992). Fragment processing in the DELPHI system. In *Proceedings of the DARPA Speech and Natural Language Workshop*, Harriman, NY, pp. 305–310.

Stede, M. (1992). The search for robustness in natural language understanding. *Artificial Intelligence Review 6*, 383–414.

Tillmann, H. G. und B. Tischer (1995). Collection and exploitation of spontaneous speech produced in negotiation dialogues. In *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, Vigsø, Denmark.

Tischer, B. (1997). Selbstkorrekturen in Dialogen: Regeln zur automatischen Syntaxverarbeitung. *Linguistische Berichte* (170), 312–344.

Tropf, H. S. (1994). Spontansprachliche syntaktische Phänomene: Analyse eines Korpus aus der Domäne "Terminabsprache". Verbmobil-Report 6, Siemens AG, München.

Wahlster, W. (1993, Juli). Verbmobil: Übersetzung von Verhandlungsdialogen. Verbmobil-Report 1, DFKI GmbH, Saarbrücken.

Wahlster, W. (1997, Juni). Verbmobil: Erkennung, Analyse, Transfer, Generierung und Synthese von Spontansprache. Verbmobil-Report 198, DFKI GmbH, Saarbrücken.

Weiss, A. (1975). *Syntax spontaner Gespräche*. Number XXXI in Sprache der Gegenwart. Schriften des Instituts für deutsche Sprache. Düsseldorf: Pädagogischer Verlag Schwann.

Winograd, T. (1983). *Language as a Cognitive Process. Volume 1: Syntax*. Reading, MA: Addison-Wesley.

Wirén, M. (1992). *Studies in Incremental Natural-Language Analysis*. Ph. D. thesis, Linköping University, Department of Computer and Information Science, Linköping, Sweden.