

ZUM AMBIGUITÄTENPROBLEM BEI DER ANALYSE KOORDINATIVER
VERBINDUNGEN

H. Hesse, A. Küstner, E. Lang

Zentralinstitut für Sprachwissenschaft AdW der DDR, Berlin

1. Bei der automatischen Sprachanalyse (speziell Syntaxanalyse) werden Texte eingegeben und automatisch analysiert. Die Aufgabe besteht darin, eine adäquate Darstellung der Strukturen zu erhalten, in der genau die r e l e v a n t e n Mehrdeutigkeiten des Materials erfasst werden. Das bedeutet

1. dass der Text weitestgehend disambiguiert wird, und
2. dass Pseudo-Mehrdeutigkeiten, wie sie etwa der Modellsatz suggerieren könnte, ausgeschlossen werden.

Art und Grad der zu analysierenden Mehrdeutigkeit einer sprachlichen Einheit (isolierte Wortform, Wortformenfolge, Satz,...) hängt dabei ab:

- a) vom Grad der Differenzierung, d.h. von der sprachlichen Ebene, die gemäss der Zwecksetzung der Analyse betrachtet wird, und
- b) von der Grösse der Umgebung, aus der Kontextinformationen auf die betrachtete Einheit bezogen werden.

Sowohl bei einer Verfeinerung der Differenzierung des Modellansatzes als auch beim Übergang von isoliert betrachteten zu komplexeren Texteinheiten während der Analyseprozedur erhöht sich im allgemeinen der Grad der Mehrdeutigkeit; es wäre zu vermuten multiplikativ. Jedoch wirkt auch eine gegenläufige Tendenz, die die Anzahl der Mehrdeutigkeiten einschränkt.

2. Für die adäquate Darstellung koordinativer Verbindungen innerhalb der Abhängigkeitsgrammatik sind Baumstrukturen unzureichend. Man kann jedoch durch spezielle Verknüpfungsregeln aus den Baumstrukturen verallgemeinerte Abhängigkeitsstrukturen ("K-Graphen") konstruieren. Einige Struktureigenschaften der K-Graphen werden skizziert. Es werden zwei unterschiedlich geartete Zerlegungsmöglichkeiten des K-Graphen gezeigt: Zur Kenntlichmachung der in einer koordinativen Verbindung enthaltenen Konjunkte dient eine Zerlegung des K-Graphen in linear geordnete Abschnitte ("Rahmengenüge", "Konjunktgefüge"); eine Zerlegung des K-Graphen in hierarchisch geordnete Strukturbestandteile macht die elementaren Struktureinheiten ersichtlich, die bei der sukzessiven Rekonstruktion der Strukturdarstellung eines Satzes im Analyseprozess benötigt werden.

3. Es wird auf das Problem von Pseudomehrdeutigkeiten eingegangen.

3.1. Die Ableitung einer mehrere gleiche Konjunktionen enthaltenden koordinativen Verbindung aus ihren Ausgangssätzen (der Koordinationsweg) ist i.a. nicht eindeutig.

Beispiel:

"Hans und Otto essen Äpfel und Birnen."

Ausgangskonjunkte:

- | | |
|------------------------|------------------------|
| 1. "Hans isst Äpfel." | 2. "Otto isst Äpfel." |
| 3. "Hans isst Birnen." | 4. "Otto isst Birnen." |

Zwei Verknüpfungswege sind möglich:

- | | |
|------------------|------------------|
| a) (1+2) + (3+4) | b) (1+3) + (2+4) |
|------------------|------------------|

Den unterschiedlichen Herleitungswegen entsprechen nicht syntaktisch relevante Ambiguitäten. Der Beispielsatz hat nur eine syntaktische Struktur. Die Herleitungswege spielen im Analyseverfahren keine Rolle; falls erwünscht, könnten sie rekonstruiert werden.

3.2. Zur Interpretation der Strukturen empirisch gegebener koordinativer Sätze ist der heuristische Ansatz einer reinen Satzverknüpfung unzureichend, da sich nicht alle koordinativen Verbindungen in grammatisch akzeptable komplette Ausgangssätze aufspalten lassen. Für eine adäquate Interpretation muss man einen Generierungsmechanismus voraussetzen, in dem zwei Prinzipien vereint sind, die kurz als Substitution und Reduktion bezeichnet werden: Eine koordinative Struktur entsteht aus einer nichtkoordinativen Ausgangsstruktur, indem gewisse Teile in ihr durch miteinander koordinierte Teile substituiert werden; die koordinative Verbindung dieser Teile kann dann durch Weglassung redundanter Teile reduziert werden. Für einen und den selben koordinativen Satz erlaubt dieser Generierungsmechanismus i.a. verschiedene Herleitungswege.

Beispiel: "Hans und Otto arbeiten."

a) Substitution des Subjekts x im Satz "x arbeiten." durch koordinierte Subjekte.
Reduktion entfällt.

b) Substitution einer nicht koordinativen Hauptsatzstruktur durch die Verbindung "Hans arbeitet und Otto arbeitet".
Reduktion der Verbindung durch Weglassung einer Verbform und Numerusveränderung in der anderen.

Derartige Unterschiede in der Herleitungsmöglichkeit rechtfertigen es jedoch nicht, von einer syntaktischen Ambiguität des Satzes zu sprechen. Der Satz hat eine syntaktisch eindeutige Struktur. Seine Herleitungswege sind syntaktisch irrelevant, bei Bedarf jedoch rekonstruierbar.

Die Fälle, in denen sich Einschränkungen für die Herleitungswege ergeben ("zwischen A und B", "Karl und Otto treffen sich."), bedürfen nicht einer gesonderten Repräsentation der zugelassenen bzw. ausgeschlossenen Herleitungswege; ihre (über die Syntax hinausgehende) Interpretation manifestiert sich

in grammatischen Beziehungen innerhalb der koordinativen Strukturen.

4. Im Modell sind die wesentlichen grammatischen Regularitäten der Koordination erfasst - als Bedingungen für die Form und für die Markierung der verallgemeinerten Abhängigkeitsstrukturen. Durch die Berücksichtigung dieser Regularitäten werden im Analyseverfahren eine Vielzahl von Ambiguitäten, die den isoliert betrachteten Wortformen eigen sind, aufgelöst. Eine weitere Disambiguierung ermöglicht der Parallelisierungseffekt, ein spezieller Wirkungseffekt der Koordination, der auf verschiedenen sprachlichen Ebenen zur Geltung kommt.

Beispiele:

- "Ich besuche Otto, und Hans besucht Max."

Die Nichtkennzeichnung der Kasuswerte im zweiten Konjunkt bewirkt Angleichung der Reihenfolge von Subjekt und Objekt an die Reihenfolge im ersten Konjunkt. Der Satz "Hans besucht Max." ist isoliert betrachtet syntaktisch zweideutig (S - P - O oder O - P - S); innerhalb der koordinativen Verbindung wird er durch die Angleichung eindeutig (S - P - O).

- "... der Besuch des Lehrers und die Einladung des Meisters..."

Die gleichartigen Genitive werden entweder beide als Subjekt- oder beide als Objektgenitiv verstanden.

- "Er raucht und sie trinkt."

Entweder sind beide Handlungen aktuell oder beide habituell.

Die unmittelbar syntaktischen Wirkungen des Parallelisierungseffektes sind ins Modell einzubeziehen. Die Voraussetzungen für eine Einbeziehung feiner differenzierter Wirkungen des Parallelisierungseffektes (semantische und pragmatische Aspekte) sind im syntaktischen Modell nicht gegeben, jedoch erleich-

tert die Kennzeichnung der im syntaktischen Sinne koordinierten Knoten eine nachträgliche Berücksichtigung derartiger Wirkungen. Dass jedoch syntaktische und semantische Parallelität auch auseinandergehen können, wird am Beispiel gezeigt.

5. Ziel einer Prozedur der automatischen syntaktischen Analyse ist es, durch die möglichst effektive Aufeinanderfolge einer Anzahl von Filtern die Bestandteile des Textes möglichst schnell zu disambiguieren. Die Mehrweganalyse sichert, dass im Endergebnis der Analyse alle Strukturen des betrachteten Satzes zur Verfügung stehen. Es werden die Prinzipien eines mehrstufigen Analyseverfahrens skizziert. Dabei wird den koordinationspezifischen Möglichkeiten zur Disambiguierung der ihnen gebührende Platz zugewiesen.