



Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

Document
D-92-04

DiTo-Datenbank
Datendokumentation zu
Verbreitung und Koordination

Judith Klein, Ludwig Dickmann

August 1992

**Deutsches Forschungszentrum für Künstliche Intelligenz
GmbH**

Postfach 20 80
D-6750 Kaiserslautern, FRG
Tel.: (+49 631) 205-3211/13
Fax: (+49 631) 205-3210

Stuhlsatzenhausweg 3
D-6600 Saarbrücken 11, FRG
Tel.: (+49 681) 302-5252
Fax: (+49 681) 302-5341

Deutsches Forschungszentrum für Künstliche Intelligenz

The German Research Center for Artificial Intelligence (Deutsches Forschungszentrum für Künstliche Intelligenz, DFKI) with sites in Kaiserslautern and Saarbrücken is a non-profit organization which was founded in 1988. The shareholder companies are Atlas Elektronik, Daimler Benz, Fraunhofer Gesellschaft, GMD, IBM, Insiders, Mannesmann-Kienzle, Philips, SEMA Group Systems, Siemens and Siemens-Nixdorf. Research projects conducted at the DFKI are funded by the German Ministry for Research and Technology, by the shareholder companies, or by other industrial contracts.

The DFKI conducts application-oriented basic research in the field of artificial intelligence and other related subfields of computer science. The overall goal is to construct *systems with technical knowledge and common sense* which - by using AI methods - implement a problem solution for a selected application area. Currently, there are the following research areas at the DFKI:

- Intelligent Engineering Systems
- Intelligent User Interfaces
- Intelligent Communication Networks
- Intelligent Cooperative Systems.

The DFKI strives at making its research results available to the scientific community. There exist many contacts to domestic and foreign research institutions, both in academy and industry. The DFKI hosts technology transfer workshops for shareholders and other interested groups in order to inform about the current state of research.

From its beginning, the DFKI has provided an attractive working environment for AI researchers from Germany and from all over the world. The goal is to have a staff of about 100 researchers at the end of the building-up phase.

Prof. Dr. Gerhard Barth
Director

DiTo Datenbank
Datendokumentation zu Verbrektion und Koordination

Judith Klein und Ludwig Dickmann

DFKI-D-92-04

Diese Arbeit wurde durch einen Forschungszuschuß (ITW 9002 0) vom Deutschen Bundesministerium für Forschung und Technologie an das DISCO Projekt am DFKI und durch das IBM Deutschland Projekt LILOG-SB, das an der Universität des Saarlandes durchgeführt wird, finanziell unterstützt.

© Deutsches Forschungszentrum für Künstliche Intelligenz 1992

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Deutsches Forschungszentrum für Künstliche Intelligenz, Kaiserslautern, Federal Republic of Germany; an acknowledgement of the authors and individual contributors to the work; all applicable portions of this copyright notice. Copying, reproducing, or republishing for any other purpose shall require a licence with payment of fee to Deutsches Forschungszentrum für Künstliche Intelligenz.

DiTo-Datenbank

Datendokumentation zu Verbrektion und Koordination *

Judith Klein[†] und Ludwig Dickmann[‡]

[†]Deutsches Forschungszentrum für Künstliche Intelligenz, GmbH
Stuhlsatzenhausweg 3, D-6600 Saarbrücken 11, FRG

Telephon: (+49 681) 302-5303, e-mail: klein@dfki.uni-sb.de

[‡]Institut für Computerlinguistik, Universität des Saarlandes
Im Stadtwald, D-6600 Saarbrücken 11, FRG

Telephon: (+49 681) 302-4138, email: dickmann@coli.uni-sb.de

Abstrakt

In dieser Arbeit werden die Daten eines Testwerkzeugs für die Fehlerdiagnose bei Syntaxkomponenten natürlichsprachlicher Systeme vorgestellt. Mit unserem Diagnostik-Tool unternehmen wir den Versuch, einen Datenkatalog zu erstellen, der die wesentlichen Phänomene deutscher Syntax erfaßt, um die Fehlerdiagnose zu unterstützen. Bisher beinhaltet der Datenkatalog die Bereiche *Verbrektion* und *Koordination*. Wir arbeiten mit anderen Gruppen¹ zusammen, die weitere Syntaxthemen entsprechend den Richtlinien unseres Ansatzes erarbeiten. Damit ausgewählte Syntaxgebiete separat abgetestet werden können, sind die Daten in einer relationalen Datenbank organisiert. In den Teildokumentationen zu den einzelnen Syntaxgebieten werden zunächst die behandelten Phänomene skizzenhaft beschrieben. Dann wird die der Datensammlung zugrundeliegende Systematik erläutert. Anschließend wird gezeigt, wie die Daten in der relationalen Datenbank organisiert sind. Die Dokumentationen enden mit einem Ausblick auf mögliche Erweiterungen der Datenbank.

*Diese Arbeit wurde durch einen Forschungszuschuß, ITW 9002 0, vom Deutschen Bundesministerium für Forschung und Technologie an das DISCO Projekt am DFKI und durch IBM Deutschland Projekt LILOG-SB, das an der Universität des Saarlandes durchgeführt wird, finanziell unterstützt.

¹IAI (Institut für angewandte Informationswissenschaft), Projekt EUROTRA in Saarbrücken und Institut für Computerlinguistik an der Universität Koblenz. Wir interessieren uns für weitere Kooperationen.

Inhaltsverzeichnis

1	Einführung	4
1.1	Einleitung	4
1.2	Organisation der Daten in der relationalen Datenbank	7
1.3	Ausblick	11
2	Daten-Dokumentation zur Verbrection	13
2.1	Einleitung	13
2.2	Die Daten zur Verbrection	14
2.2.1	Ausgangsdaten und Systematik	14
2.2.2	Einschränkungen	18
2.2.3	Behandelte Rektionsbereiche	20
2.2.4	Ungrammatische Sätze und ihre Klassifizierung	21
2.3	Organisation der Daten in der Datenbank	22
2.3.1	Das Datenschema	22
2.3.2	Fehlertypen	24
2.3.3	Beispielanfragen	25
2.4	Ausblick	27

A	Liste der verwendeten Verben	28
B	Liste der Komplement-Kombinationen	31
3	Daten-Dokumentation zur Satzkoordination	33
3.1	Einleitung	33
3.2	Die Daten zur Koordination	34
3.2.1	Ausgangsdaten und Systematik:	34
3.2.2	Einschränkungen	35
3.2.3	Behandelte Koordinations-Phänomene	36
3.2.4	Ungrammatische Sätze und ihre Klassifizierung	38
3.2.5	Generalisierungen	39
3.3	Organisation der Daten in der Datenbank	40
3.3.1	Das Datenschema	40
3.3.2	Fehlertypen	42
3.3.3	Beispielanfragen	47
3.4	Ausblick	48
A	Liste der möglichen Kombinationen von Tilgungen und Permutationen bei Verb-Zweit-Stellung	49
B	Liste der möglichen Kombinationen von Tilgungen und Permutationen bei Verb-Letzt-Stellung	52
	Literaturverzeichnis	54

Kapitel 1

Einführung

Ludwig Dickmann und Judith Klein

1.1 Einleitung

Die vorliegende Arbeit beschreibt die Daten, die die Grundlage des Diagnostik-Tools DiTo bilden.¹ Diesem einführenden Kapitel folgen die Datendokumentationen zu *Verbrektion* und *Satzkoordination*. Die Dokumentationen sind so weit wie möglich parallel aufgebaut, um

- DiTo-Benutzern die Orientierung zu erleichtern
- Kooperationspartnern, die an Erweiterungen zu DiTo arbeiten, eine Art Standard für die Erstellung von Dokumentationen vorzugeben.

DiTo stellt den Versuch dar, einen linguistischen Datenkatalog zu erstellen, der möglichst alle wichtigen Bereiche der deutschen Syntax anhand von Beispieldaten systematisch abdecken soll. Die Beispieldaten und die syntaktischen Annotationen sind in einer relationalen Datenbank organisiert, um

¹Eine genaue Beschreibung der technischen Details der Datenbank-Implementierung liefert Diagne 92.

- den Zugriff auf die Daten zu erleichtern,
- die Einträge konsistent zu halten,
- gute Voraussetzungen für die Erweiterung um neue Syntaxbereiche bereitzustellen und
- verschiedene logische Sichten auf die Daten zu gewähren.

Die Beispielsätze sind aus technischen Gründen in der Datenbank ohne Kommas, Umlaute und Großbuchstaben eingetragen.

Das Vokabular, mit dem die Beispielsätze konstruiert sind, beschränkt sich so weit wie möglich auf den Bereich Personalverwaltung. Um den Umfang des Vokabulars möglichst klein zu halten, wird im Rahmen einer Datenbankerweiterung unter anderem ein Utility entwickelt, das dem Benutzer bei Bedarf alle verwendeten Wörter auflistet.

Systematik der ungrammatischen Sätze

Um bei dem angestrebten Einsatz dieser Testmenge zu erreichen, daß

- alle und nur die richtigen Sätze analysiert werden und
- sinnvolle Toleranzgrade bei der Fehlerbehandlung herausgearbeitet werden können,

umfaßt die Datensammlung eine Auflistung ungrammatischer Sätze, d.h. Sätze, die vorgegebene syntaktische Forderungen verletzen. Diese Sätze sind mit * bzw. ? für Grenzfälle gekennzeichnet. Dabei haben wir versucht, die für die den jeweiligen syntaktischen Bereich wahrscheinlichen Fehlerquellen herauszufiltern und zu klassifizieren. Auf dieser Grundlage wurden zu jedem Fehlertyp systematisch Beispiele konstruiert.

Eingeschränkte Parameter

Damit möglichst klar definierte und somit kontrollierbare Daten als Testmenge vorliegen, sind für die in dieser Dokumentation beschriebenen Syntaxgebiete folgende Beschränkungen zweckmäßig.

- **Syntaktische Funktion der Satzelemente**

Die Komplemente wurden nicht nach ihrer syntaktischen Funktion im Satz (Subjekt, direktes Objekt, indirektes Objekt, ...) bezeichnet, sondern nach morphologischen Kriterien (NOM, DAT, ACC, ZU-INF, ...), so daß eine weitgehend theorieunabhängige Beschreibung möglich ist. Durch diese Theorieunabhängigkeit ist gewährleistet, daß mit DiTo Systeme getestet werden können, die auf verschiedenen Syntaxtheorien basieren.

- **Kontextunabhängigkeit**

Die Beispielsätze sollen kontextunabhängig wohlgeformt sein, d.h. mögliche Beziehungen zu anderen Satzstrukturen und Satzinhalten werden nicht berücksichtigt.

- **Determination**

Bei NPs beschränken wir uns soweit wie möglich darauf, den bestimmten Artikel zu verwenden, um eine eindeutige Kasusmarkierung zu gewährleisten.

- **Kasusmarkierung**

Die nominalen Satzelemente wurden soweit wie möglich durch nicht ambige Kasusmarkierung kenntlich gemacht. Morphologische Ambiguitäten zwischen Feminin in Nominativ und Akkusativ sowie Neutrum in Nominativ und Akkusativ schließen wir dadurch aus, daß wir diese Formen nicht verwenden.

- **Besonderheiten innerhalb der Satzelemente**

Syntaktische Besonderheiten innerhalb der Satzelemente wie z.B. die Unterscheidung zwischen *Raising* und *Equi* innerhalb der Infinitivkomplemente bei der Verbrektion werden nicht berücksichtigt, wenn sie in keinem direkten Bezug zu dem illustrierten Syntaxbereich stehen.

1.2 Organisation der Daten in der relationalen Datenbank

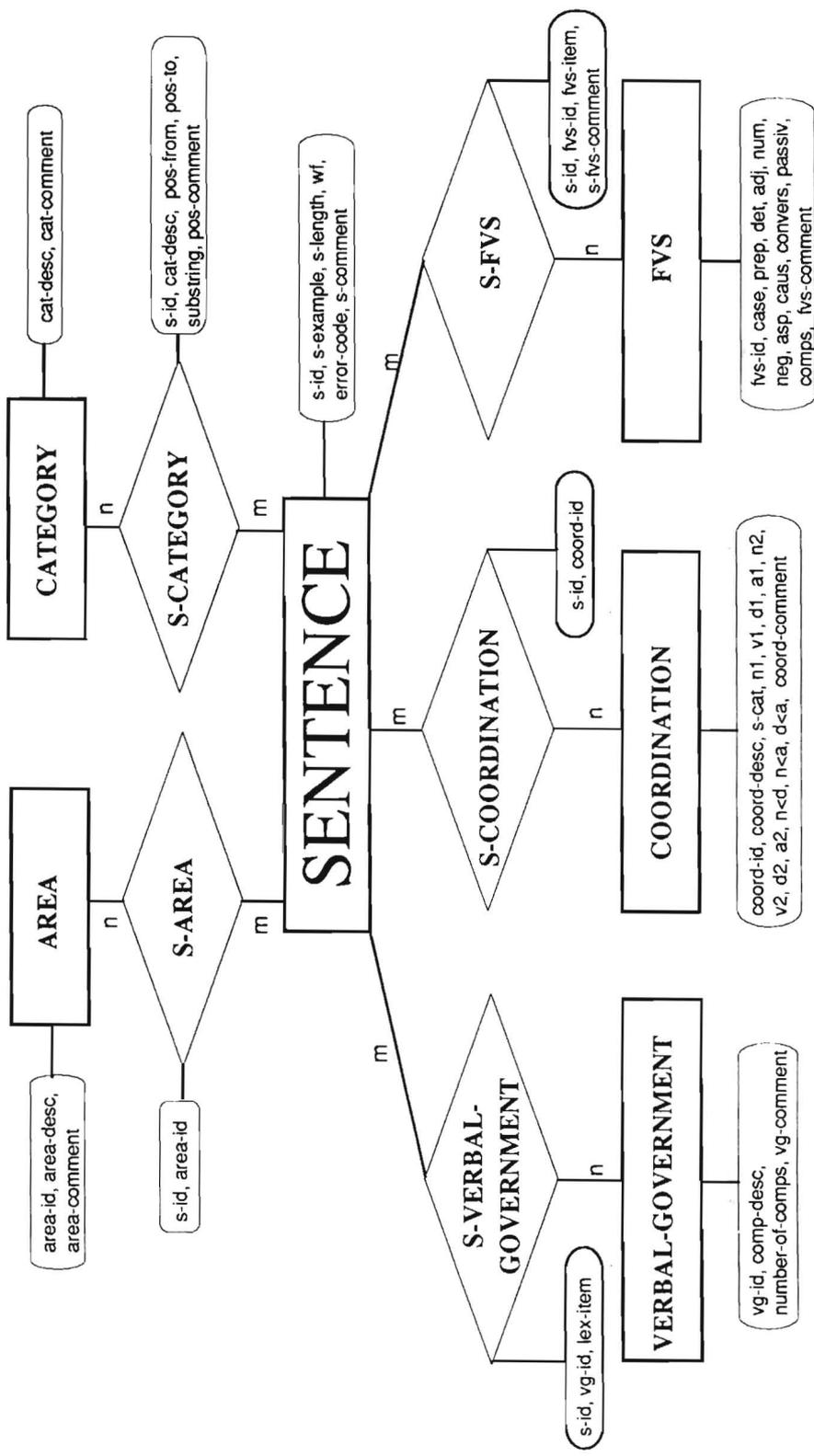
Damit letztlich nicht nur eine Liste von Beispielsätzen vorliegt, die syntaktische Phänomene illustrieren, sondern Sätze, die nach wesentlichen syntaktischen Eigenschaften untersucht werden können, sind die Daten in folgender Weise in der Datenbank strukturiert:

Das Datenschema

In der zentralen Relation SENTENCE stehen die Attribute, die die Sätze nach allgemeinen Kriterien charakterisieren.

- **Satznummer** (s-id)
Jedem Satz ist eine eindeutige Nummer zugewiesen.
- **Beispielsatz** (s-example)
In diesem Feld stehen die Satzbeispiele.
- **Satzlänge** (s-length)
Diese Information dient unter anderem dazu, bestimmte Konsistenzprüfungen durchzuführen.
- **Wohlgeformtheit** (wf)
Die Werte des Attributes *wf* geben die Grammatikalität des Satzes an.
- **Fehlerklassifikation** (error-code)
Den Fehlertypen² der Syntaxphänomene sind bisher Nummern von -1 bis -13 zugewiesen.
- **Kommentarfeld** (s-comment)
Dieses Attribut bietet die Möglichkeit, Besonderheiten eines Satzes zu notieren. Handelt es sich um einen ungrammatischen Satz, steht in diesem Feld die Erklärung des Fehlertyps.

²Siehe S.24 Verbrenktion und S.42-45 Koordination.



Das Entity-Relationship-Modell der Datenbank

Die Einträge haben folgende Form:

s-id	s-example	s-length	wf	error-code	s-comment
1138	der vorschlag bringt den studenten darauf dass der plan falsch ist.	11	1	0	null
3076	der professor schenkte den blumenstraus und verkaufte den roman dem kommilitonen.	11	1	0	null
3021	*der professor schenkte der sekretaerin einen blumenstraus und der student verkaufte dem kommilitonen.	13	0	- 10	rechtstilgung

Abbildung 1.1: Beispiel für die Einträge in der Relation SENTENCE

Der Eintrag 3021 ist ein Beispiel für einen ungrammatischen Satz. Also hat das Feld *Wohlgeformtheit* den Wert 0 und im Feld *Fehlerklassifikation* steht der entsprechende Fehlercode. Das *Kommentarfeld* gibt in diesem Fall die Aufschlüsselung des Fehlertyps.

Zusätzlich zu diesen Attributen sind in der Relation S-CATEGORY allgemeinsyntaktische Informationen bezüglich der Satzstruktur erfasst. Es geht dabei um die Position relevanter Kategorien im Satz. Die ausgewählten Kategorien sind Entitäten des Typs CATEGORY.

- **Kategoriekürzel** (cat-desc)
Hier stehen die Bezeichnungen der Kategorien.
- **Kategorienamen** (cat-comment)
Die Werte sind die jeweiligen Kategorienamen.

Die Relation S-CATEGORY umfaßt folgende Attribute:

- **Satznummer** (s-id)
Über die Satznummer sind die einzelnen Satzelemente eindeutig dem entsprechenden Ausgangssatz zugeordnet.
- **Kategoriekürzel** (cat-desc)
siehe oben

- **Anfangsposition der Kategorie** (pos-from)
Der numerische Wert gibt an, wo die betreffende Kategorie im Satz beginnt.
- **Schlußposition der Kategorie** (pos-to)
Der numerische Wert gibt an, wo die betreffende Kategorie im Satz endet. Deshalb muß pos-to immer größer oder gleich pos-from sein.
- **Satzelemente** (substring)
In diesem Feld stehen das Wort oder die Wörter aus dem Ausgangssatz.
- **Kommentarfeld** (pos-comment)
Dieses Attribut bietet die Möglichkeit, Besonderheiten bzgl. allgemeinsyntaktischer Informationen festzuhalten.

Jeder grammatische Satz ist partiell in seine Teilkategorien aufgesplittet. Für den Satz mit der Nummer 1138 in Abbildung 1.2 sieht der Eintrag in der Relation S-CATEGORY wie folgt aus:

s-id	cat-desc	pos-from	pos-to	substring	pos-comment
1138	np	1	2	der vorschlag	----
1138	f_verb	3	3	bringt	----
1138	np	4	5	den studenten	----
1138	cor	6	6	darauf	----
1138	np	8	9	der plan	----
1138	sc-comp	7	11	dass der plan falsch ist	----

Abbildung 1.2: Datensatz zur Beschreibung der absoluten Position der Kategorien im Satz

In der Relation AREA wird den syntaktischen Phänomenen eine Identifikationsnummer zugewiesen. Folgende Attribute stehen in AREA:

- **Bereichskennung** (area-id)
1 steht für Verbrektion, 2 für Koordination und 3 für Funktionsverbgefüge.

- **Syntaxphänomen** (area-desc)
Hier steht der Name des Syntaxbereiches.
- **Kommentarfeld** (area-comment)
Dieses Feld bietet die Möglichkeit, das Phänomen kurz zu spezifizieren.

Die Werte in S-AREA zeigen an, welchem syntaktischen Phänomen der jeweilige Beispielsatz zugeordnet ist. Die Attribute sind:

- **Satznummer** (s-id)
- **Bereichskennung** (area-id)
- **Kommentarfeld** (area-comment)

Datenbank-Anfragen

Die Daten können unter zwei verschiedenen Gesichtspunkten erfragt werden:

- Fragen nach Beispielsätzen, die bestimmte syntaktisch-formale Eigenschaften haben
- Fragen nach der syntaktisch-formalen Beschreibung eines konkreten Beispielsatzes

In den Dokumentationen zu den einzelnen Syntaxgebieten stehen einige Beispielanfragen mit ihren Ergebnissen.

1.3 Ausblick

Bei der Implementierung von DiTo wurde ausschließlich *public domain*-Software verwendet, da wir DiTo allen interessierten Gruppen zur Verfügung stellen möchten, sei es zur Benutzung oder zur Weiterentwicklung. Zur Zeit arbeitet Brigitte Krenn am IAI in Saarbrücken an Daten zu *Funktionsverbgefügen*, die auch in die Datenbank eingetragen und entsprechend dokumentiert werden sollen. Martin Volk vom Institut für Computerlinguistik an der Universität Koblenz hat vorgeschlagen, Relativsätze zu behandeln. Für weitere Kooperationsangebote sind wir dankbar.

In Vorbereitung befindet sich eine *Media Track* Demo-Version zu DiTo, die eine Beispielsitzung mit Datenbank-Anfragen und Ergebnissen enthalten soll.

An Interface-Software zur Anbindung von DiTo an zu testende Systeme sowie weiteren Utilities (z.B. zur Konsistenzprüfung) arbeiten wir zur Zeit noch.

Kapitel 2

Daten-Dokumentation zur Verbrektion

Judith Klein

2.1 Einleitung

Die Datensammlung beschreibt das syntaktische Phänomen der Verbrektion im Deutschen. Wir haben diesen Bereich als einen der ersten ausgewählt, weil er als ein zentrales Thema innerhalb der deutschen Syntax anzusehen ist. Viele andere Syntaxgebiete haben fast immer in irgendeiner Form mit den Eigenschaften des regierenden Verbs eines Satzes zu tun und können dann durch Querverweise an die Verbrektion angebunden werden.

Das Ziel dieses Datenkatalogs ist die möglichst vollständige Zusammenstellung der Kombinationen obligatorischer Komplemente, die die deutschen Satzchemata bilden. Da wir in dieser Arbeit eine *syntaktische* Beschreibung der Verbrektion versuchen, sind die regierten Elemente nicht als semantische Argumente der ausgewählten Verben beschrieben. Es geht um die minimale Satzstruktur, so daß fakultative Komplemente und Adjunkte, d.h. modifizierende Adjektive, Adverbien, genitivische Nominalphrasen und Partikeln nicht vorkommen. Ein Test, der den Status der Komplemente zeigt, ist eine besondere Form der Extraposition, wie

folgende Beispiele aus der Satzsammlung zeigen:

- (1) Der Manager arbeitet, und zwar lange.
- (2) *Der Vortrag dauert, und zwar lange.

Die ungrammatischen Sätze zeigen, daß hier die extraponierten Elemente obligatorisch sind.

Das Kapitel über die Verbrektion ist wie folgt aufgebaut: In Abschnitt 2.2 wird kurz umrissen, welche Daten hier zusammengetragen sind und nach welcher Systematik sie strukturiert wurden. Anschließend wird die Organisation der Daten in der Datenbank beschrieben und an einigen Beispielanfragen veranschaulicht. Zum Abschluß führen wir mögliche Erweiterungen im Bereich *Verbrektion* an.

2.2 Die Daten zur Verbrektion

Die syntaktischen Merkmale der Verbrektion werden anhand von 212 grammatischen Beispielsätzen annähernd exhaustiv dargestellt.

2.2.1 Ausgangsdaten und Systematik

In den Arbeiten zur Verbrektion sind die von den Verben regierten Elemente unterschiedlich benannt und z.T. unterschiedlichen Klassen zugeteilt.¹ Wir haben die Komplemente soweit wie möglich nach morpho-syntaktischen Merkmalen klassifiziert. Die Bezeichnungen der Komplemente innerhalb der Komplement-Kombinationen ergeben sich aus dem jeweiligen Komplementnamen.

Die vorkommenden Komplemente sind wie folgt bezeichnet:

NOM : nominativische Nominalphrase (**NOM**inative NP)
GEN : genitivische Nominalphrase (**GEN**itive NP)

¹Engel/Schumacher 76 geben einen guten Überblick über die Bezeichnungen einzelner Komplemente bei verschiedenen Autoren.

DAT	: dativische Nominalphrase (DAT ive NP)
ACC	: akkusativische Nominalphrase (ACC usative NP)
P-NOM	: Präpositionalphrase mit Nominativ-NP (P- NOM inative)
P-DAT	: Präpositionalphrase mit Dativ-NP (P- DAT ive)
P-ACC	: Präpositionalphrase mit Akkusativ-NP (P- ACC usative)
PRED-NO	: Prädikatives Nomen (PRED icative NO un)
PRED-ADJ	: Prädikatives Adjektiv (PRED icative ADJ ective)
PP-DIR	: Directionale Präpositionalphrase
PP-LOC	: Locative Präpositionalphrase
AD-DUR	: Duratives AD jektiv oder AD verbial
ADJ-MOD	: Modales ADJ ektiv
PRED-P	: Prädikativphrase (PRED icative P hrase)
SC-daß	: Nebensatz mit daß eingeleitet (S ubordinate C lause)
SC-ob	: Nebensatz mit ob eingeleitet (S ubordinate C lause)
SC-w	: Nebensatz mit w- Wort eingeleitet (S ubordinate C lause)
INF	: reiner INF initiv
zu-INF	: zu-INF initiv
COR	: Korrelat (COR relative) (Präpositionaladverb oder <i>es</i>)

Besonderheiten einiger Komplemente:

- Innerhalb der *präpositionalen* Komplemente sind zwei Klassen zu unterscheiden:
 1. *Präpositionalobjekte* mit Nominalphrasen in Nominativ, Dativ und Akkusativ
 2. Präpositionale *Angaben*

Die beiden Komplementarten unterscheiden sich im Status der Präposition, mit der sie stehen. Im ersten Fall handelt es sich um Präpositionalobjekte im traditionellen Sinn, deren Präposition semantisch leer ist und nur die Funktion hat, der abhängigen NP den Kasus zuzuweisen. Die Form der Präposition hängt vom regierenden Verb ab. Diese Klasse umfaßt folgende Komplementarten: P-NOM, P-ACC, P-DAT. Verben wie *stimmen gegen*, *grauen vor* verlangen diese Komplementart.

Bei der zweiten Gruppe der Präpositionalphrasen trägt die Präposition noch eigene Bedeutung und wird vom regierenden Verb in Form einer semantischen Klasse oder Relation selegiert. Diese präpositionalen Komplemente sind innerhalb der Komplemente PP-DIR, PP-LOC, AD-DUR, PRED-P realisiert und werden von Verben wie *blicken auf*, *legen auf* regiert.

- Unter die Komplement-Bezeichnung *PRED-P (Prädikativphrase)* fallen Ausdrücke wie die in dem Satz "Der Manager tut **wie der Präsident.**" Es war schwierig, diese Komplemente einzuordnen, so daß wir uns für diese neutrale Bezeichnung entschieden haben.
- Bei einigen Komplementbeschreibungen mit Satzkomplementen und zu-Infinitiven tritt als Besonderheit ein Komplement auf, das mit *COR* für Korrelat bezeichnet ist. Es handelt sich hierbei um Präpositionaladverbien (*darauf, davon, darum, ...*) und um *es* in expletiver Funktion, wobei wir das expletive *es* in Topikposition, d.h. vor dem Verb in Verb-Zweit-Stellung ignorieren. Über den Status des Korrelats wird in der zitierten Literatur keine klare Aussage getroffen. In dieser Datensammlung ist es in den Sätzen, in denen es vorkommt, obligatorisches Komplement.
- Aus den Arbeiten zum Thema Verbrektion wird nicht deutlich, ob die Verben in den unten aufgeführten Beispielen das Komplement *Modalangabe (ADJ-MOD)* subkategorisieren, oder ob es sich bei den gekennzeichneten Elementen jeweils um das Präfix der trennbaren Verben *guttun* bzw. *leicht fallen* handelt.

(3) Der Urlaub tut dem Präsidenten *gut*. (nom_dat_adj-mod)

(4) Der Test fällt dem Studenten *leicht*. (nom_dat_adj-mod)

Einerseits können diese Elemente modifiziert werden (*besonders gut, sehr leicht*), was für die Behandlung als Modalangabe spricht, andererseits sind sie nicht - wie bei Modalangaben üblich - frei austauschbar. Wir haben uns in unserer Datensammlung dafür entschieden, sie zu den Modalangaben zu rechnen. In einer Zusatzinformation weisen wir allerdings auf den idiomatischen Charakter dieser Komplementart hin.

- Bei den Infinitivkonstruktionen sind die Beispielsätze der Art

(5) Der Manager hört den Studenten kommen. (nom_acc_inf)

nicht nach der Small-Clause-Theorie analysiert. Die Komplementbeschreibung *nom_acc_inf* zeichnet die NP *den Studenten* und die VP *kommen* als separate Komplemente aus, wodurch in der Analyse eine größere Oberflächennähe erreicht wird. Innerhalb der Small-Clause-Theorie erhielte man weniger detaillierte Information, nämlich nur, daß es sich um ein Infinitivkomplement handelt, innerhalb dessen eine NP vorkommt. Über die morphosyntaktischen Merkmale dieser NP wäre nichts Näheres ausgesagt.

Systematik der Komplement-Kombinationen:

Jede Komplement-Kombination wird durch mindestens einen Beispielsatz verdeutlicht. Sind für eine Kombination mehrere Verben aufgeführt, so liegt es

entweder daran, daß diese Verben mit verschiedenen Präpositionen stehen oder daß besonders interessante Fälle dabei sind oder aber das Verb ist sowieso schon im Lexikon vorhanden und hat zu dem bereits aufgenommenen Rektionsrahmen noch zusätzliche. Wir haben die aus der Literatur zusammengetragenen Verb-Valenzrahmen und formalen Satzstrukturbeschreibungen nach folgender Systematik zusammengestellt:

- **Kombinationen nicht-verbaler Komplemente**

Diese Klasse umfaßt reine nominale Komplemente, präpositionale Komplemente, prädikative und adjektivische Komplemente.

1. Reine nominale Komplemente:

(6) Der Manager enthebt den Präsidenten des Amtes.
(nom_acc_gen)

2. Nominale Komplemente kombiniert mit präpositionalen Komplementen:

(7) Der Vorschlag bringt den Studenten auf den Lösungsweg. (nom_acc_p_acc)

(8) Der Manager wohnt auf dem Land. (nom_pp_loc)

3. Nominale Komplemente kombiniert mit prädikativen Komplementen:

(9) Der Manager ist der Präsident. (nom_pred_no)

(10) Der Manager ist tolerant. (nom_pred_adj)

4. Nominale Komplemente kombiniert mit adjektivischen Komplementen:

(11) Der Manager wohnt gut. (nom_adj_mod)

- **Kombinationen sententialer Komplemente**

In dieser Gruppe sind Komplemente der ersten Klasse mit Nebensätzen kombiniert, die mit *daß*, *ob* oder einem 'w' - Fragewort eingeleitet sind.

1. Nebensätze, die mit *daß* eingeleitet sind:

(12) Dem Präsidenten entgeht, daß der Manager arbeitet.
(dat_sc_daß)

(13) Der Manager besteht darauf, daß der Student arbeitet.
(nom_cor_sc_daß)

2. Nebensätze, die mit *ob* eingeleitet sind:

(14) Der Manager fragt, ob der Student arbeitet. (nom_sc_ob)

(15) Der Manager beurteilt den Studenten danach, ob er kompetent ist. (nom_acc_cor_sc_ob)

3. Nebensätze, die mit einem 'w' - Wort eingeleitet sind:

(16) Den Präsidenten interessiert, wen der Manager einstellt.
(acc_sc-w)

(17) Der Manager achtet darauf, wer arbeitet. (nom_cor_sc-w)

• Kombinationen infinitivischer Komplemente

Die Infinitivkomplemente sind eingeteilt in Konstruktionen mit reinem Infinitiv und solche mit zu-Infinitiv. Beide Infinitivformen sind mit verschiedenen Arten nicht-verbaler Komplemente kombiniert.

1. reiner Infinitiv:

(18) Der Manager hört den Studenten kommen. (nom_acc_inf)

2. zu-Infinitiv:

(19) Dem Studenten imponiert, den Präsidenten zu sehen.
(dat_zu-inf)

(20) Der Vorschlag bringt den Studenten darauf, den Vortrag zu überarbeiten. (nom_acc_cor_zu-inf)

In den Kombinationsklassen verbaler Komplemente zeigt jeweils das zweite Beispiel eine Kombination mit Korrelat.

2.2.2 Einschränkungen

Die eingetragenen Verben sollen jeweils für einen Rektionsrahmen prototypisch sein, d.h. wir haben nicht alle deutschen Verben in die einzelnen Klassen eingeordnet, sondern mindestens ein Verb für jede Komplement-Kombination ausgewählt. Ausgegangen sind wir von der Zusammenstellung von Helbig² und haben dann die Verben und ihre Valenzrahmen anhand anderer Arbeiten aus diesem Bereich kontrolliert und ergänzt.

Zu den in Kapitel 1 beschriebenen Parametern, die eingeschränkt wurden, um möglichst kontrollierbare Daten zu erhalten, sind noch einige hinzuzufügen, die allein den Bereich Verbrektion betreffen.

• Kontraktion von Präposition und definitivem Artikel

Bei den präpositionalen Komplementen gibt es keine Kontraktion der Präposition mit dem nachfolgenden Artikel. (*im* : *in dem*)

²Helbig 80 hat die Verben semantisch nach Wortfeldern klassifiziert.

- **Wortstellung**

Da es um Rektionsabhängigkeiten geht, bei denen das Vorkommen der Komplemente und nicht ihre Position im Satz im Vordergrund steht, haben wir die Beispielsätze in *unmarkierter Wortstellung* konstruiert.

- **Fakultative Modifikation**

Innerhalb der Beispielsätze stehen nur die für das syntaktische Gebiet obligatorischen Elemente. Adjunkte jeglicher Art sind in den Sätzen nicht vorhanden.

- **Verbform**

Im Matrixsatz und in den Satzkomplementen steht jeweils nur ein verbaler Kopf, d.h. Auxiliärverben, die Tempus kennzeichnen, sind ausgeschlossen. Modalverben, die einen Infinitiv regieren, wurden in die Datensammlung aufgenommen. Die Verbform ist festgelegt auf 3. Person Singular Präsens Indikativ Aktiv.

Ausnahmen zu den allgemeinen Richtlinien

Einige Beispielsätze entsprechen nicht ganz den festgelegten Restriktionen. In diesen Fällen wurde auf die strikte Einhaltung verzichtet, weil die Sätze sonst unnatürlich wären. Diese Beispiele zeigen relevante Rektionsrahmen und wurden deshalb in die Datensammlung aufgenommen. Eine zusätzliche Information weist auf die Besonderheit in diesen Sätzen hin.

- (21) Der Schrank dient als *Stütze*.
- (22) Der Manager betrachtet es als *Aufgabe*, daß der Präsident den Vortrag hält.
- (23) Der Manager betrachtet es als *Aufgabe*, den Vertrag zu verlängern.
- (24) Den Präsidenten verlangt nach *Ruhe*.
- (25) Der Manager legt *Wert* darauf, daß der Student arbeitet.
- (26) Der Manager legt *Wert* darauf, den Vortrag zu halten.
- (27) Der Manager wird *Präsident*.
- (28) Der Manager bezeichnet den Studenten als *Idioten*.

- (29) Der Manager will nach *Saarbrücken*.
- (30) Der Manager fährt den Präsidenten nach *Saarbrücken*.
- (31) Der Manager wohnt in *Saarbrücken*.
- (32) *Peter* bleibt der Präsident.
- (33) *Peter* wird wie der Präsident.
- (34) Der Manager heißt Peter.
- (35) Der Manager schimpft den Präsidenten *einen* Idioten.
- (36) Der Manager hat *ein* Projekt laufen.
- (37) *Keines* Beweises bedarf, daß der Manager arbeitet.
- (38) Der Manager verzeiht dem Studenten, daß *er* kündigt.
- (39) Der Manager zwingt den Studenten dazu, daß *er* arbeitet.
- (40) Der Manager hindert den Studenten daran, daß *er* arbeitet.
- (41) Der Manager interessiert den Studenten dafür, daß *er* den Vortrag hält.

In all diesen Fällen entspricht ein nominales Element nicht der gewünschten Form.³ Die ersten 11 Sätze haben eine NP ohne Artikel. In den Beispielen (32) - (34) steht ein Eigename. Beispiel (35) und (36) haben keinen definiten, sondern einen indefiniten Artikel. Satz (37) hat statt eines definiten Artikels ein Negationselement und in den letzten Beispielsätzen steht ein Personalpronomen.

2.2.3 Behandelte Rektionsbereiche

Bisher sind in dieser Datensammlung noch keine reflexiven Verben und keine Verben mit abtrennbarem Präfix enthalten. Eine Liste der aufgenommenen Verben ist in Anhang A1 zusammengestellt. Die Komplement-Kombinationen stehen in

³Vgl. Kapitel 1, S.6.

Anhang A2. Funktionsverbgefüge, die bezüglich der Rektionseigenschaften der Funktionsverben auch zum Thema Verbrektion zählen, sind über einen Querverweis an die Verbrektionsdaten angeschlossen. Die besonderen Eigenschaften dieses Phänomens werden separat behandelt.⁴

2.2.4 Ungrammatische Sätze und ihre Klassifizierung

Zu jeder Komplement-Kombination gibt es eine Zusammenstellung ungrammatischer Sätze. Diese Negativbeispiele zeigen drei Fehlerarten, die die Rektionsabhängigkeiten betreffen. Wir haben uns für diese drei Arten entschieden, auch wenn sicherlich noch andere syntaktischen Aspekte berücksichtigt werden könnten. Insgesamt sind 225 ungrammatische Sätze in die Datensammlung aufgenommen worden. Die einzelnen Fehlertypen sind:

- **Es fehlt ein obligatorisches Komplement**

Ein beliebig ausgewähltes Komplement, das in der Komplementbeschreibung gefordert ist, ist nicht im Beispielsatz vorhanden. Bei den nicht-verbalen Komplementen handelt es sich dabei meist um nicht-nominativische Nominalphrasen. Diese Fehlerart fehlt bei Komplement-Kombinationen mit nur einem obligatorischen Komplement, weil hier nur noch das regierende Verb übrigbleiben würde, wie die Beispielsätze (44) und (45) zeigen:

(42) Der Manager beschuldigt *den Studenten*
des Diebstahls. (nom_acc_gen)

(43) *Der Manager beschuldigt des Diebstahls.

(44) *Der Manager* kommt. (nom)

(45) *Kommt.

- **Es steht ein Komplement zuviel**

Der ungrammatische Beispielsatz enthält ein Komplement mehr als in der Komplementbeschreibung erlaubt ist. Diese zusätzlichen Komplemente sind so gewählt, daß sie nicht als Adjunkte interpretiert werden können. Außerdem haben wir versucht, jeweils Komplemente zu wählen, die, wenn sie allein in dem Satz auftreten, in den Valenzrahmen des regierenden Verbes passen.

⁴Vgl. Krenn 92.

- (46) Der Manager beschuldigt den Studenten des Diebstahls.
- (47) *Der Manager beschuldigt den Studenten des Diebstahls, daß der Student den Präsidenten ärgert .

- **Es steht ein falsches Komplement**

Die syntaktische Form eines Komplements des ungrammatischen Beispielsatzes entspricht nicht der in der Komplementbeschreibung geforderten.

- (48) Der Manager beschuldigt den Studenten des Diebstahls.
- (49) *Der Manager beschuldigt den Studenten *den Diebstahl*.

2.3 Organisation der Daten in der Datenbank

2.3.1 Das Datenschema

In der Relation VERBAL-GOVERNMENT stehen die syntaktischen Merkmale, die für die Verbrektion kennzeichnend sind:

- **vg-id** (Identifikationsnummer der Rektionskombination)
Die Identifikationsnummer weist jeder Merkmalskombination eine eindeutige Nummer zu.
- **comp-desc** (Komplementbeschreibung)
Innerhalb der Komplementbeschreibung entspricht die Reihenfolge der Komplemente dem Vorkommen der Elemente im Beispielsatz bei unmarkierter Wortstellung.⁵
- **number-of-comps** (Anzahl der Komplemente)
Die Datensammlung enthält Beispielsätze mit 1-4 Komplementen.
- **vg-comment** (Kommentarfeld)
Das Kommentarfeld bietet die Möglichkeit, besondere Eigenschaften einzelner Komplemente einzutragen.

⁵Eine vollständige Liste der Komplement-Kombinationen und ihrer Identifikationsnummer steht in Anhang B.

Folgende Tabelle illustriert die Einträge in VERBAL-GOVERNMENT:

vg-id	comp-desc	number - of - comps	vg-comment
008	nom_dat_acc	3	null
055	dat_sc-ob	2	null

Abbildung 2.1: Beispiel für die Einträge der Verbreaktionseigenschaften

Über *vg-id* = 008 sind die syntaktischen Merkmale z.B. mit dem Satz *Der Manager gibt dem Studenten den Computer* verknüpft. Die Werte der *vg-id* = 055 treffen z.B. auf den Satz *Dem Präsidenten entfällt, ob der Student kommt* zu.

Über die Relation S-VERBAL-GOVERNMENT sind die Beispielsätze mit den syntaktischen Merkmalen verknüpft. Diese Zuordnung wird durch die Kombination von *s-id* und *vg-id* erreicht.

- **Satznummer** (*s-id*)
- **Identifikationsnummer der Rektionskombination** (*vg-id*)
- **Verbeintrag** (*lex-item*)
Die Werte von *lex-item* sind die vorkommenden regierenden Verben in Infinitivform. Bei Verben, die für ein Präpositionalobjekt subkategorisieren, steht in *lex-item* nur die Verbform. Die Präposition wird als zum Komplement zugehörig betrachtet.

Die Einträge in S-VERBAL-GOVERNMENT sehen wie folgt aus:

s-id	vg-id	lex-item
1022	008	geben
1165	055	sagen

Abbildung 2.2: Beispiel für die Einträge in S-VERBAL-GOVERNMENT

2.3.2 Fehlertypen

Die ungrammatischen Beispielsätze der Verbrektion illustrieren - wie auf Seite 20 erläutert - drei Fehlertypen. Diese stehen in der Relation SENTENCE im Feld *error-code* mit den Werten -1 bis -3. Im Kommentarfeld (*s-comment*) steht bei den ungrammatischen Sätzen die Kurzbeschreibung des Fehlertyps.

Folgendes Beispiel soll diese Systematik illustrieren. Der grammatische Ausgangssatz mit seiner Komplement-Beschreibung lautet:

(50) Der Vorschlag bringt den Studenten darauf, daß der Plan falsch ist.
(nom_acc_cor_sc-daß)

- **es fehlt ein Komplement: -1**

(51) *Der Vorschlag bringt den Studenten, daß der Plan falsch ist.

- **es steht ein Komplement zuviel: -2**

(52) *Der Vorschlag bringt den Studenten darauf auf den Lösungsweg, daß der Plan falsch ist.

- **es steht ein falsches Komplement: -3**

(53) *Der Vorschlag bringt *dem Studenten* darauf, daß der Plan falsch ist.

2.3.3 Beispielanfragen

In den ersten 4 Anfragen werden Beispielsätze erfragt, die bestimmte syntaktische Merkmale haben. Die Umsetzung der natürlichsprachlichen Fragestellung in die Datenbanksprache ist kursiv gedruckt. Wie das Ergebnis der jeweiligen Anfrage aussieht, zeigen die nachfolgenden Daten.

- Zeige die Satznummer und die dazugehörigen grammatischen Sätze aus dem Bereich Verbrektion, die als Komplemente **mindestens** ein Dativ-Objekt und einen zu-Infinitiv haben!

retrieve s-id s-example where match(comp-desc, dat) and match(comp-desc, zu-inf) and wf= 1 and area-id = 1

1198 dem studenten imponiert den praesidenten zu sehen.
1199 dem studenten gefaellt zu programmieren.
1200 dem studenten liegt zu programmieren.
1201 dem studenten graut davor den praesidenten zu sehen.
1202 dem studenten liegt daran den praesidenten zu sehen.
1214 der manager untersagt dem studenten zu programmieren.
1215 der manager ueberlaesst es dem studenten den vortrag zu ueberarbeiten.

- Zeige die Satznummern und die dazugehörigen grammatischen Sätze aus dem Bereich Verbrektion, die als Komplemente **ausschließlich** ein Dativ-Objekt und einen zu-Infinitiv haben!

retrieve s-id s-example where comp-desc = "dat_zu-inf" and wf = 1 and area-id = 1

1198 dem studenten imponiert den praesidenten zu sehen.
1199 dem studenten gefaellt zu programmieren.
1200 dem studenten liegt zu programmieren.

- Zeige die ungrammatischen Sätze und die jeweilige Fehlerbezeichnung aus dem Bereich Verbrektion mit der Komplementbeschreibung: nom_acc_cor_sc_dass!

retrieve s-example s-comment where comp-desc = "nom_acc_cor_sc_dass" and wf = 0

*der vorschlag bringt den studenten, dass der plan falsch ist.
es fehlt ein Komplement

* der vorschlag bringt den studenten auf den loesungsweg darauf, dass der plan falsch ist.

es steht ein Komplement zuviel

* der vorschlag bringt dem studenten darauf, dass der plan falsch ist.

es steht ein falsches Komplement

- Zeige die Komplementbeschreibung und alle Sätze aus dem Bereich Verbktion, die 4 Komplemente haben, wobei eins der Komplemente ein daß-Satz ist!

retrieve comp-desc s-example where number-of-comps = 4 and match(comp-desc, sc-dass) and area-id = 1

nom_acc_cor_sc-dass der vorschlag bringt den studenten darauf dass der plan falsch ist.

nom_acc_cor_sc-dass der manager legt wert darauf dass der student arbeitet.

nom_acc_cor_sc-dass *der vorschlag bringt den studenten, dass der plan falsch ist.

...

nom_pred_p_cor_sc-dass der manager betrachtet es als aufgabe dass der praesident den vortrag haelt.

nom_pred_p_cor_sc-dass *der manager betrachtet es, dass der praesident den vortrag haelt.

...

Das letzte Beispiel zeigt eine Anfrage, die, von einem konkreten Satz ausgehend, syntaktische Merkmale dieses Satzes erfragt.⁶

- Zeige die Position der NPs in den Sätzen mit der Komplemente-Beschreibung **nom_dat_acc**.

retrieve s-id cat-desc cat-position substring where cat-desc = "np" and comp-desc = "nom_dat_acc"

1022 np 1 2 der manager

1022 np 4 5 dem studenten

1022 np 6 7 den computer

1023 np 1 2 der manager

1023 np 4 5 dem praesidenten

...

⁶Bei Anfragen nach der Position der Kategorien im Satz haben wir die Möglichkeit, statt *pos-from* und *pos-to* das Attribut *cat-position* zu verwenden, das die Ergebniswerte von *pos-from* und *pos-to* kombiniert.

2.4 Ausblick

Durch die Organisation der Daten in der relationalen Datenbank ist es leicht, die Datensammlung um weitere Einträge zu erweitern. Folgende Ergänzungen könnten interessant sein:

- Erarbeitung der Reflexivverben und der Verben mit abtrennbarem Präfix. Möglicherweise müßten ein oder zwei Merkmale zu der Relation Verbal-Government hinzugefügt werden, damit die Besonderheiten dieser Verben erfaßt werden. Diese Ergänzung bereitet aber wegen des modularen Charakters der Datenbankarchitektur keine Schwierigkeiten.
- Zuordnung möglichst aller deutschen Verben zu der erarbeiteten Systematik
- Zuordnung aller in der Datenbank enthaltenen Verben zu all ihren Rektionsrahmen. Bisher sind die Verben in der Regel *einer* ihrer Komplement-Kombinationen zugeordnet. Um die Abdeckung zu vervollständigen, könnten Beispielsätze konstruiert werden, die die Verben mit allen ihren Komplement-Kombinationen erfassen.
- Komplement-Rahmen und Beispielsätze mit fakultativen Komplementen

Anhang A

Liste der verwendeten Verben

achten
ärgern
bedürfen
befestigen
beginnen
begrüssen
behaupten
berücksichtigen
beschuldigen
bestehen
betrachten
betrauen
beurteilen
bewahren
bezeichnen
bitten
bleiben
blicken
bringen
dauern
denken
dienen
dürfen
eilen
entfallen
entgehen
entheben
erstaunen

fahren
folgen
fragen
freuen
frieren
geben
gedenken
gefallen
gehören
gelten
glauben
grauen
haben
hängen
halten
heissen
hindern
hören
hoffen
imponieren
interessieren
können
kommen
lassen
legen
lehren
liegen
liegen
lügen
machen
mögen
muessen
nehmen
nennen
sagen
schimpfen
schliessen
schwindeln
sehen
sein
sollen
spielen
sprechen
stecken
stehen

stimmen
streben
streiten
tun
überlassen
übernachten
überraschen
untersagen
verdächtigen
verdanken
verlangen
vermuten
vertrauen
verzeihen
wählen
wagen
werden
wissen
wohnen
wollen
ziehen
zwingen

Anhang B

Liste der Komplement-Kombinationen

001: nom
002: acc
003: dat
004: nom_acc
005: nom_dat
006: nom_gen
007: nom_acc_acc
008: nom_dat_acc
009: nom_acc_gen
010: nom_p-nom
011: nom_p-acc
012: nom_p-dat
013: acc_p-dat
014: dat_p-dat
015: nom_acc_p-acc
016: nom_acc_p-dat
017: nom_pp-dir
018: nom_pp-loc
019: nom_ad-dur
020: nom_adj-mod
021: nom_pred-p
022: dat_adj-mod
023: nom_acc_pp-dir
024: nom_acc_pp-loc
025: nom_acc_ad-dur
026: nom_acc_adj-mod

027: nom_acc_pred-p
028: nom_dat_adj-mod
029: nom_pred-adj
030: nom_pred-no
031: sc-dass
032: nom_sc-dass
033: acc_sc-dass
034: dat_sc-dass
035: gen_sc-dass
036: p-dat_sc-dass
037: adj-mod_sc-dass
038: nom_dat_sc-dass
039: nom_p-dat_sc-dass
040: acc_mp_sc-dass
041: nom_cor_sc-dass
042: dat_cor_sc-dass
043: nom_acc_cor_sc-dass
044: nom_pred-p_cor_sc-dass
045: nom_sc-w
046: acc_sc-w
047: dat_sc-w
048: nom_dat_sc-w
049: nom_p-dat_sc-w
050: nom_cor_sc-w
051: nom_acc_cor_sc-w
052: nom_sc-ob
053: acc_sc-ob
054: dat_sc-ob
055: nom_dat_sc-ob
056: nom_cor_sc-ob
057: nom_acc_cor_sc-ob
058: nom_inf
059: nom_adj-mod_inf
060: nom_acc_inf
061: nom_zu-inf
062: acc_zu-inf
063: dat_zu-inf
064: nom_acc_zu-inf
065: nom_dat_zu-inf
066: nom_cor_zu-inf
067: dat_cor_zu-inf
068: acc_adj-mod_zu-inf
069: nom_acc_cor_zu-inf
070: nom_dat_cor_zu-inf
071: nom_pred-p_cor_zu-inf

Kapitel 3

Daten-Dokumentation zur Satzkoordination

Ludwig Dickmann

3.1 Einleitung

Dieses Kapitel beschreibt die Sammlung systematisch erstellter Beispieldaten zu bestimmten Phänomenen der Satzkoordination.¹ In Punkt 3.2 werden zuerst die behandelten Phänomene anhand von Beispielen kurz skizziert. Anschließend wird die den Daten zugrunde liegende Systematik erläutert. Danach werden einige interessante Generalisierungen aufgeführt, die sich aus den gesammelten Daten ableiten lassen. Punkt 3.3 behandelt die Organisation der Daten in der Datenbank, wobei die einzelnen Datenfelder beschrieben werden. Es folgen eine kommentierte Liste der Fehlertypen für ungrammatische Sätze sowie mehrere Beispielanfragen. Unter Punkt 3.4 schließlich folgt ein Ausblick auf mögliche Erweiterungen von DiTo im Bereich der Koordination.

¹Der Verfasser dankt Bettina Rehse für Diskussion und Anregungen zur Behandlung der Koordination in der theoretischen Linguistik.

3.2 Die Daten zur Koordination

3.2.1 Ausgangsdaten und Systematik:

Ausgangspunkt der Datensammlung waren zwei komplexe Sätze, die aus jeweils zwei mit *und* verknüpften Teilsätzen bestanden. Diese Teilsätze enthielten jeweils ein dreistelliges Verb (*schicken* bzw. *verkaufen*), und zwar (a) in Verb-Letzt-Stellung bzw. (b) in Verb-Zweit-Stellung:

a) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß schenkte, und der Student dem Kommilitonen den Roman verkaufte.

b) Der Professor schenkte der Sekretärin den Blumenstrauß und der Student verkaufte dem Kommilitonen den Roman.

Von diesen Sätzen wurden systematisch alle Kombinationen von Permutationen und Tilgungen in beiden Konjunkten betrachtet.

Für Verb-Zweit-Sätze gibt es insgesamt 41 mögliche Kombinationen von Tilgungen und Permutationen.² Also gibt es für den komplexen Satz $41 * 41 = 1681$ Sätze.

Für Verb-Letzt-Sätze gibt es insgesamt 32 mögliche Kombinationen von Tilgungen und Permutationen, und somit für den komplexen Satz $32 * 32 = 1024$ Sätze.

Insgesamt gibt es für Verb-Zweit- und Verb-Letzt-Sätze $1681 + 1024 = 2705$ Sätze. Diese sehr hohe Zahl wurde durch die unter 3.2.2. aufgeführten Einschränkungen auf insgesamt 534 Sätze reduziert. Davon entfallen auf die Verb-Zweit-Stellung 297 Sätze (56 grammatisch, 241 ungrammatisch) und auf die Verb-Letzt-Stellung 237 Sätze (52 grammatisch, 185 ungrammatisch).

Bei der Erstellung der Daten waren vor allem die **Vollständigkeit** und die **Systematik** entscheidend. Auf die Anwendung einer bestimmten linguistischen Theorie kam es uns nicht an.

²Eine genaue Auflistung der möglichen Kombinationen für Verb-Zweit- und Verb-Letzt-Sätze steht im Anhang.

3.2.2 Einschränkungen

- Um einer kombinatorischen Explosion vorzubeugen, wurden nur Permutationen von NPs im zweiten Konjunkt betrachtet.
- Es wurden keine Sätze betrachtet, in denen in beiden Konjunkten dieselbe Konstituente (z.B. die NP im Nominativ) getilgt wurde, da diese Sätze zwangsläufig ungrammatisch sind.
- Es wurden nur Sätze betrachtet, in denen insgesamt maximal vier Tilgungen auftreten. Treten mehr Tilgungen auf, dann ist eine Konstituente in beiden Konjunkten getilgt (vgl. vorhergehender Punkt). Bei vier Tilgungen ist jede Konstituente in einem der beiden Konjunkte realisiert, und obwohl diese Sätze alle falsch sind, sind sie dennoch interessant (vgl. 3.2.4).
- Alle Sätze wurden nur in ihrer Oberflächenstruktur³ betrachtet, da in vielen Fällen Sätze mit verschiedener Tiefenstruktur eine Oberflächenstruktur teilen, wie folgendes Beispiel zeigt:

Für den Satz

(54) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß schenkte, und der Student verkaufte.

gibt es verschiedene Tiefenstrukturen, beispielsweise

(55) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß schenkte, und der Student **O**dat **O**akk verkaufte.

(56) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß schenkte, und **O**akk der Student **O**dat verkaufte.

Da in der Oberflächenstruktur nicht zu erkennen ist, auf welche Tiefenstruktur der Satz zurückgeht, wurde der Satz nur einmal aufgenommen.

Aufgrund der Erfassung von reinen Oberflächenstrukturen bleiben bei der Koordination von Sätzen mit Verb-Zweit-Stellung mehr Sätze übrig: Bei der Verb-Zweit-Stellung steht das Hauptverb nicht am Satzende, sondern "im" Satz. Dadurch ergibt sich eine größere Anzahl an Oberflächenstrukturen, denn wenn im zweiten Konjunkt beispielsweise nur das Hauptverb und eine NP realisiert sind, kann die NP sowohl **vor** als auch **nach** dem Hauptverb stehen:

³Die Begriffe Tiefenstruktur und Oberflächenstruktur werden hier nicht in einem theoriespezifischen Sinn benutzt, sondern dienen nur zur Unterscheidung zwischen den Ausgangssätzen (Tiefenstruktur), und den durch Anwendung von Permutation(en) und Tilgung(en) entstandenen und in die Datenbank aufgenommenen Sätzen (Oberflächenstruktur).

- (57) *Der Professor schenkte der Sekretärin den Blumenstrauß und *der Student* verkaufte.
- (58) *Der Professor schenkte der Sekretärin den Blumenstrauß und verkaufte *der Student*.

3.2.3 Behandelte Koordinations-Phänomene

Durch die unter 3.2.1. näher erläuterte systematische Betrachtung aller Tilgungs- und Permutationsmöglichkeiten decken sowohl die Verb-Letzt- als auch die Verb-Zweit-Beispielsätze Phänomene wie Gapping, Right-Node-Raising und Links- und Rechtstilgung ab.⁴

- Beispielsätze für Gapping:

Unter Gapping versteht man die Tilgung eines Hauptverbs und einer oder mehrerer NPs innerhalb eines Satzes:

- (59) Der Professor schenkte der Sekretärin den Blumenstrauß, und **der Student verkaufte dem Kommilitonen** den Roman.
- (60) Der Professor schenkte der Sekretärin den Blumenstrauß, und **der Student** den Roman.⁵
- (61) Ich glaube, daß der Professor der Sekretärin **den Blumenstrauß schenkte**, und der Student dem Kommilitonen den Roman verkaufte.
- (62) Ich glaube, daß der Professor der Sekretärin, und der Student **der Kommilitonen** den Roman verkaufte.

- Beispielsätze für Right-Node-Raising:

Unter Right-Node-Raising versteht man die Tilgung einer rechtsstehenden *Konstituenten* im linken Konjunkt, zu der im rechten Konjunkt eine strukturell identische Konstituenten existiert:

⁴Gapping, Right-Node-Raising und Links- und Rechtstilgung sind keine Datenbank-Kategorien, sondern stellen abstraktere Beschreibungen dar. Für einen Überblick über diese Phänomene siehe z.B. van Oirsouw 1985.

⁵Der eigentliche Ausgangssatz für diesen Beispielsatz ist natürlich "Der Professor schenkte der Sekretärin den Blumenstrauß, und der Student schenkte der Sekretärin den Roman." Da es uns bei dieser Datensammlung aber nur um die syntaktische Struktur der Beispielsätze geht, haben wir obenstehenden Satz als Ausgangssatz aufgeführt. Dies gilt analog auch für die Beispielsätze zu Right-Node-Raising, Rechtstilgung und Linkstilgung.

- (63) Der Professor schenkte der Sekretärin **den Blumenstrauß**, und der Student verkaufte dem Kommilitonen den Roman.
- (64) Der Professor schenkte der Sekretärin, und der Student verkaufte dem Kommilitonen den Roman.
- (65) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß **schenkte**, und der Student dem Kommilitonen den Roman verkaufte.
- (66) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß, und der Student dem Kommilitonen den Roman verkaufte.

- Beispielsätze für Rechtstilgung im ersten Konjunkt:

Unter Rechtstilgung versteht man die Tilgung strukturell identischen rechtsstehenden *Materials*, also, im Gegensatz zu Right-Node-Raising, auch *mehrerer Konstituenten*:

- (67) Der Professor schenkte **der Sekretärin den Blumenstrauß** , und der Student verkaufte dem Kommilitonen den Roman.
- (68) Der Professor schenkte, und der Student verkaufte dem Kommilitonen den Roman.
- (69) Ich glaube, daß der Professor der Sekretärin **den Blumenstrauß schenkte** , und der Student dem Kommilitonen den Roman verkaufte.
- (70) Ich glaube, daß der Professor der Sekretärin, und der Student dem Kommilitonen den Roman verkaufte.

- Beispielsätze für Linkstilgung im zweiten Konjunkt:

Unter Linkstilgung versteht man die Tilgung strukturell identischen linksstehenden *Materials*:

- (71) Der Professor schenkte der Sekretärin den Blumenstrauß und **der Student** verkaufte dem Kommilitonen den Roman.
- (72) Der Professor schenkte der Sekretärin den Blumenstrauß und verkaufte dem Kommilitonen den Roman.

- (73) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß schenkte und **der Student dem Kommilitonen** den Roman verkaufte.
- (74) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß schenkte und den Roman verkaufte.

Zu beachten ist, daß nur die Satzkoordination betrachtet wurde, nicht die Konstituentenkoordination im allgemeinen.⁶ Durch die systematische Erstellung der Sätze erhielten wir jedoch auch Beispielsätze, die man als Konstituentenkoordination analysieren könnte. Ein Satz wie

- (75) Der Professor schenkte der Sekretärin den Blumenstrauß und den Roman.

scheint zwar ein Beispiel für NP-Koordination zu sein, tritt hier aber doch auf, weil er gut in ein allgemeines Schema der Satzkoordination paßt, wie z.B.:

- (76) Der Professor schenkte der Sekretärin den Blumenstrauß und **O_{nom}**
O_v **O_{dat}** den Roman.

3.2.4 Ungrammatische Sätze und ihre Klassifizierung

In der Datensammlung sind auch ungrammatische Sätze enthalten. Durch die oben aufgeführten Einschränkungen der Satzauswahl sind allerdings nur "interessant" ungrammatische Sätze enthalten, wie z.B.

- (77) * Ich glaube, daß einen Blumenstrauß schenkte, und der Student dem Kommilitonen.
- (78) * Ich glaube, daß der Professor einen Blumenstrauß schenkte, und der Student dem Kommilitonen.

Diese Sätze enthalten, verteilt auf die beiden Konjunkte, alles, was **ein** korrekter Teilsatz benötigt, nämlich NPs in Nominativ, Dativ und Akkusativ sowie ein finites Hauptverb und stellen somit eine echte Bewährungsprobe für ein zu testendes System dar.

⁶Beispiele für noch nicht behandelte Koordinationsphänomene sind unter Punkt 3.4. "Ausblick" aufgeführt.

3.2.5 Generalisierungen

Obwohl es *nicht* Ziel dieser Arbeit war, Generalisierungen zur Koordination zu erschließen, ist es doch interessant zu sehen, daß durch die zugrundeliegende Systematik bei der Erzeugung und Auswahl der Sätze Generalisierungen über die (Un-)Grammatikalität von Sätzen möglich sind. Die folgenden Beispiele führen wir weniger aus Gründen der Dokumentation auf, sondern vielmehr, um zu zeigen, welche Nebenergebnisse eine solche Datensammlung erbringen kann.

- Sätze sind ungrammatisch, wenn im ersten Konjunkt das Subjekt fehlt:
 - (79) **Der Professor** schenkte der Sekretärin den Blumenstrauß und der Student verkaufte dem Kommilitonen den Roman.
 - (80) * Schenkte der Sekretärin den Blumenstrauß und der Student verkaufte dem Kommilitonen den Roman.

Eine Ausnahme stellen die Sätze dar, in denen durch weitere Tilgungen und durch Permutationen im zweiten Konjunkt dann Konstituentenkoordination vorliegt:

- (81) **Der Sekretärin und dem Kommilitonen** verkaufte der Student den Roman.
- Sätze sind ungrammatisch, wenn bei Verb-Zweit-Stellung im zweiten Konjunkt die Vorfeldkonstituente getilgt ist, und durch Permutation eine Nachfeld-NP die Vorfeld-Position einnimmt:
 - (82) Der Professor schenkte der Sekretärin den Blumenstrauß und **der Student** verkaufte **dem Kommilitonen** den Roman.
 - (83) * Der Professor schenkte der Sekretärin den Blumenstrauß und **dem Kommilitonen** verkaufte den Roman.
- Korrekte Sätze werden ungrammatisch, wenn durch Permutation(en) ein Symmetriebruch vorliegt:
 - (84) Der Professor schenkte der Sekretärin und der Student dem Kommilitonen **den Roman**.
 - (85) * Der Professor schenkte der Sekretärin und der Student **den Roman** dem Kommilitonen.

Im zweiten Konjunkt des zweiten Satzes hat sich die Dativ-NP durch Permutation zwischen die Nominativ-NP und die Akkusativ-NP geschoben. Dadurch haben die zwei Konjunkte nun verschiedene Strukturen, und der Satz ist ungrammatisch.

3.3 Organisation der Daten in der Datenbank

3.3.1 Das Datenschema

In der Relation COORDINATION stehen die syntaktischen Merkmale, die für Koordination kennzeichnend sind. Diese Merkmale werden im Datenschema durch folgende Attribute repräsentiert:

- **coord-id** (Identifikation des koordinierten Satzes):

Diese Identifikationsnummer dient zur Kennzeichnung von Kombinationen bestimmter Koordinations-Eigenschaften (vgl. nächster Punkt, "coord-desc").

- **coord-desc** (Koordinationsbeschreibung):

Beschreibung der Konstituentenstruktur im koordinierten Satz. Dabei steht n1 für die Nominativ-NP im ersten Konjunkt, v1 für das Hauptverb im ersten Konjunkt, d1 für die Dativ-NP im ersten Konjunkt, usw.

- **s-cat** (Satzkategorie):

Für die Satzategorie gibt es zwei mögliche Werte: "V-second" bzw. "V-last" für Verb-Zweit- bzw. Verb-Letzt-Sätze.

- **n1, v1, ..., a2:**

Auch in diesen Feldern steht n1 für die Nominativ-NP im ersten Konjunkt, v1 für das Hauptverb im ersten Konjunkt, usw.

Diese Felder dienen der Beschreibung der einzelnen Konstituenten.

Sie haben drei mögliche Werte:

- 0: Konstituente nicht getilgt, nicht permutiert
- 1: Konstituente nicht getilgt, aber permutiert⁷
- 2: Konstituente getilgt

Mit diesen Angaben läßt sich beispielsweise erfassen, ob im zweiten Konjunkt die NPs in Dativ und Akkusativ miteinander permutiert sind. Dies ist der Fall, wenn $d2 = 1$, $a2 = 1$ und $n2 = 0$.

⁷Dieser Wert kann nur bei n2, d2 und a2 auftreten, da die beiden Hauptverben sowieso nicht permutiert werden können, und da wir uns bei den NP-Permutationen auf die NPs im zweiten Konjunkt beschränkt haben (vgl. Punkt 3.2.2. "Einschränkungen".)

- **n-d, n-a und d-a:**

In diesen Feldern steht *n* für die Nominativ-NP, *d* für die Dativ-NP und *a* für die Akkusativ-NP. Diese Felder dienen der Angabe der linearen Präzedenz der NPs im zweiten Konjunkt. So bedeutet “n-a”, daß im zweiten Konjunkt die Nominativ-NP vor der Akkusativ-NP steht.

Sie haben drei mögliche Werte:

- 0: Lineare Präzedenz verletzt, da eine oder zwei der betroffenen NPs permutiert sind.
- 1: Lineare Präzedenz gegeben.
- null: keine Angabe zur linearen Präzedenz möglich, da eine oder zwei der betroffenen NPs getilgt sind.

Alle möglichen Permutationen der NPs im zweiten Konjunkt lassen sich einer Kombination dieser LP-Angaben eindeutig zuordnen:

NP-Folge	n-d	n-a	d-a
N D A	1	1	1
N A D	1	1	0
D N A	0	1	1
D A N	0	0	1
N D A	1	0	0
A D N	0	0	0

Folgende Tabelle zeigt die Datenbankeinträge in der Relation COORDINATION für den Satz

- (86) * Der Professor schenkte den Blumenstrauß und dem Kommilitonen der Student den Roman.

coord-id	coord-desc	s-cat	n1	v1	d1	a1	n2	v2	d2	a2	n-d	n-d	n-d	coord - comment
3081	N1_V1_A1_ D2_N2_A2_	V-second	0	0	2	0	1	2	1	0	0	1	1	null

Abbildung 3.1: Beispiel für die Einträge in der Relation COORDINATION

Die Relation S-COORDINATION dient zur Verknüpfung von SENTENCE und COORDINATION. Sie enthält die Attribute *s-id* und *coord-id*. Im Gegensatz zu *vg-id* (vgl. Teil 2) wird durch *coord-id* nur jeweils **ein** Satz beschrieben. Deshalb ist *coord-id* mit *s-id* identisch.

3.3.2 Fehlertypen

Die Fehlertypen, die wir verwendet haben, folgen keiner bestimmten linguistischen Theorie und erheben auch keinen Anspruch auf Vollständigkeit. In Bezug auf ungrammatische Sätze stellen sie lediglich eine von mehreren Möglichkeiten dar. Die Fehlercodes können zur Orientierung in unseren Daten dienen und bei der Formulierung von Suchanfragen nützlich sein.

Über die Grammatikalitätsbewertung könnte man bei manchen Sätzen natürlich diskutieren. Dies fällt vor allem bei Sätzen mit Permutationen auf, da bei diesen Sätzen Unterschiede bei der Akzeptanz von Wortstellungsvarianten auftreten.

In der folgenden Auflistung der Fehlertypen ist neben Beispielen und Kurzbeschreibungen auch angegeben, wie die einzelnen Fehlercodes in der Datenbank eingetragen sind. Für die Koordination werden zur Zeit die Fehlercodes von -4 bis -13 benutzt.

- **vier Tilgungen**

In s-comment eingetragen als "vier-tilgungen", in error-code als "-4".

(87) * Der Professor schenkte, und dem Kommilitonen den Roman.

- **Wortstellung**

s-comment: "wortstellung", error-code: "-5".

Zwei Sätze enthalten die gleichen Konstituenten, unterscheiden sich aber in ihrer Anordnung der NPs. Manche dieser Wortstellungsvarianten sind grammatisch, andere nicht:

(88) Ich glaube, daß der Professor den Blumenstrauß, und der Student **den Roman dem Kommilitonen** verkaufte.

(89) * Ich glaube, daß der Professor den Blumenstrauß, und der Student **dem Kommilitonen den Roman** verkaufte.

- **Gapping im ersten Konjunkt bei Verb-Zweit-Stellung**

s-comment: “gapping”, error-code: “-6”.

Gapping ist im koordinierten Satz nur im zweiten Konjunkt möglich, im ersten nicht.

- (90) Der Professor **schenkte der Sekretärin** den Blumenstrauß und der Student **verkaufte dem Kommilitonen** den Roman.
- (91) Der Professor schenkte der Sekretärin den Blumenstrauß und der Student den Roman.
- (92) * Der Professor den Blumenstrauß und der Student verkaufte dem Kommilitonen den Roman.

- **Linkstilgung im ersten Konjunkt**

s-comment: “linkstilgung”, error-code: “-7”.

Linkstilgung ist im koordinierten Satz nur im zweiten Konjunkt möglich, im ersten nicht.

- (93) Ich glaube, daß **der Professor der Sekretärin** den Blumenstrauß schenkte, und **der Student dem Kommilitonen** den Roman verkaufte.
- (94) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß schenkte und den Roman verkaufte.
- (95) * Ich glaube, daß den Blumenstrauß schenkte, und der Student dem Kommilitonen den Roman verkaufte.

- **unzulässige Tilgung einer Konstituenten im ersten Konjunkt**

s-comment: “mitteltilgung-1”, error-code: “-8”.

Es liegt kein Gapping vor, und, im Gegensatz zu Rechts- und Linkstilgung sowie Right-Node-Raising, wird hier eine Konstituente getilgt, die im ersten Konjunkt nicht am Rand, sondern in der Mitte des Satzes steht:

- (96) Ich glaube, daß der Professor **der Sekretärin** den Blumenstrauß schenkte, und der Student dem Kommilitonen den Roman verkaufte.
- (97) * Ich glaube, daß der Professor den Blumenstrauß schenkte, und der Student dem Kommilitonen den Roman verkaufte.

- (98) Der Professor schenkte **der Sekretärin** den Blumenstrauß, und der Student verkaufte dem Kommilitonen den Roman.
- (99) * Der Professor schenkte den Blumenstrauß, und der Student verkaufte dem Kommilitonen den Roman.

- **unzulässige Tilgung einer Konstituenten im zweiten Konjunkt**

s-comment: “mitteltilgung-2”, error-code: “-9”.

Auch hier wird eine Konstituente getilgt, die in der Mitte des Satzes steht, nur diesmal im zweiten Konjunkt.

- (100) Ich glaube, daß der Professor der Sekretärin den Blumenstrauß schenkte, und der Student **dem Kommilitonen** den Roman verkaufte.
- (101) * Ich glaube, daß der Professor der Sekretärin den Blumenstrauß schenkte, und der Student den Roman verkaufte.
- (102) Der Professor schenkte der Sekretärin den Blumenstrauß, und der Student verkaufte **dem Kommilitonen** den Roman.
- (103) * Der Professor schenkte der Sekretärin den Blumenstrauß, und der Student verkaufte den Roman.

- **Rechtstilgung im zweiten Konjunkt:**

s-comment: “rechtstilgung”, error-code: “-10”.

Rechtstilgung ist im koordinierten Satz nur im ersten Konjunkt möglich, im zweiten nicht.

- (104) Der Professor schenkte der Sekretärin **den Blumenstrauß**, und der Student verkaufte dem Kommilitonen **den Roman**.
- (105) Der Professor schenkte der Sekretärin, und der Student verkaufte dem Kommilitonen den Roman.
- (106) * Der Professor schenkte der Sekretärin den Blumenstrauß, und der Student verkaufte dem Kommilitonen.

- **Im zweiten Konjunkt bei Verb-Zweit-Stellung: Vorfeldkonstituente getilgt und Neubesetzung des Vorfelds durch andere Konstituente:**

s-comment: “neues-vorfeld”, error-code: “-11”.

Durch Linkstilgung erhält man zunächst einen korrekten Satz, der dann durch die Permutation einer Nachfeld-NP an die freigewordene Vorfeldposition ungrammatisch wird.

- (107) Der Professor schenkte der Sekretärin den Blumenstrauß und **der Student** verkaufte **dem Kommilitonen** den Roman.
- (108) Der Professor schenkte der Sekretärin den Blumenstrauß und verkaufte **dem Kommilitonen** den Roman.
- (109) * Der Professor schenkte der Sekretärin den Blumenstrauß und **dem Kommilitonen** verkaufte den Roman.

- **Linkstilgung und Rechtstilgung in einem Konjunkt**

s-comment: “links-rechts-tilgung”, error-code: “-12”.

In einem einzelnen Konjunkt des koordinierten Satzes können Rechts- und Linkstilgung nicht gemeinsam auftreten.

- (110) Der Professor schenkte der Sekretärin den Blumenstrauß und **der Student** verkaufte dem Kommilitonen **den Roman**.
- (111) * Der Professor schenkte der Sekretärin den Blumenstrauß und verkaufte dem Kommilitonen.

In einem **koordinierten Satz** können Links- und Rechtstilgung gemeinsam auftreten, nämlich Rechtstilgung im ersten Konjunkt und Linkstilgung im zweiten:

- (112) Der Professor schenkte der Sekretärin **den Blumenstrauß** und **der Student** verkaufte dem Kommilitonen den Roman.
- (113) Der Professor schenkte der Sekretärin und verkaufte dem Kommilitonen den Roman.

- **Doppelte Rechtstilgung im ersten Konjunkt**

s-comment: “doppelte-rechtstilgung”, error-code: “-13”.

- (114) Der Professor schenkte **der Sekretärin den Blumenstrauß**, und der Student dem Kommilitonen den Roman.

Die **einfache** Rechtstilgung im ersten Konjunkt liefert einen grammatischen Satz:

- (115) Der Professor schenkte **der Sekretärin**, und der Student dem Kommilitonen den Roman.

Die **doppelte** Rechtstilgung im ersten Konjunkt nicht mehr:

- (116) * Der Professor schenkte, und der Student dem Kommilitonen den Roman.

Rechtstilgung im ersten Konjunkt liefert i.A. einen grammatischen Satz (vgl. Punkt 3.2.3. "Behandelte Koordinations-Phänomene".) Auch die Sätze mit dem s-comment "doppelte-rechtstilgung" wären korrekt, wenn im zweiten Konjunkt das Hauptverb nicht getilgt wäre:

- (117) Der Professor schenkte, und der Student **verkaufte** dem Kommilitonen den Roman.

Es handelt sich hierbei also um einen absoluten Spezialfall der Rechtstilgung, für den wir die Bezeichnung "doppelte Rechtstilgung" gewählt haben, um einen intuitiv verständlichen Namen für dieses Phänomen zu haben, nicht weil wir glauben, daß dieser Name eine gute linguistische Erklärung darstellt.

Bei manchen Sätzen kann die Ungrammatikalität auf verschiedene Fehler zurückgeführt werden, wie folgendes Beispiel zeigt:

- (118) * Der Professor schenkte den Blumenstrauß, und der Student dem Kommilitonen den Roman.

Hier liegen zwei mögliche Fehler vor, nämlich:

a) Fehler 5: Mitteltilgung im ersten Konjunkt. Der entsprechende korrekte Satz wäre in diesem Fall:

- (119) Der Professor schenkte **der Sekretärin** den Blumenstrauß, und der Student dem Kommilitonen den Roman.

b) Fehler 2: Wortstellung. Der entsprechende korrekte Satz wäre in diesem Fall:

- (120) Der Professor schenkte den Blumenstrauß, und der Student **den Roman dem Kommilitonen**.

In der Datenbank eingetragen ist Fehler 5, da Mitteltilgung präziser erfaßbar als Wortstellung ist und somit den vom linguistischen Standpunkt interessanteren Fehler darstellt. Solche Sonderfälle sind also nicht prinzipiell auf eine einzige Erklärung zurückzuführen.

3.3.3 Beispielanfragen

Wie in den Beispielanfragen zur Verbrektion werden auch hier in den ersten vier Anfragen Beispielsätze erfragt, die bestimmte syntaktische Merkmale haben.

- Zeige die grammatischen Sätze aus dem Bereich Koordination mit ihrer Beschreibung der Konstituentenstruktur, in denen im zweiten Konjunkt Gapping (Tilgung des Hauptverbs und einer oder mehrerer NPs) auftritt:

retrieve s-example coord-desc where wf = 1 and area-id = 2 and v2 = 2 and (n2 = 2 or d2 = 2 or a2 = 2)

der professor schenkte der sekretärin den blumenstraus und dem kommilitonen den roman.

N1_V1_D1_A1_D2_A2

(sowie 17 weitere Sätze)

- Zeige die grammatischen Sätze mit ihrer Satznummer und mit ihrer Beschreibung der Konstituentenstruktur, in denen bei Verb-Zweit-Stellung im zweiten Konjunkt die NP im Nominativ mit der NP im Akkusativ vertauscht ist:

retrieve s-id s-example coord-desc where wf = 1 and area-id = 2 and s-cat= "V-second" and n2 = 1 and a2 = 1 and d2 = 0

3006 der professor schenkte der sekretärin den blumenstraus und den roman verkaufte dem Kommilitonen der student.

N1_V1_D1_A1_A2_V2_D2_N2

(sowie 6 weitere Sätze)

- Zeige die ungrammatischen Sätze mit ihrer Satznummer, in denen "doppelte Rechtstilgung" als Fehler auftritt:

retrieve s-id s-example where s-comment = "doppelte-rechtstilgung"

3223 *der professor schenkte und der student dem kommilitonen den roman.

(sowie 5 weitere Sätze)

- Zeige die grammatischen Sätze mit ihrer Satznummer und mit ihrer Beschreibung der Konstituentenstruktur, in denen im zweiten Konjunkt die Nominativ-NP vor der Dativ-NP steht und die Akkusativ-NP getilgt ist:

retrieve coord-desc s-id s-example where n-d = 1 and a2 = 2 and wf = 1 and area-id = 2

N1_V1_D1_A1_A2_N2_D2 3033 der professor schenkte der sekretaerin den blumenstraus und der student dem kommilitonen.

N1_D1_A1_V1_N2_D2 ich glaube dass der professor der sekretaerin den blumenstraus schenkte und der student dem kommilitonen.

Das letzte Beispiel zeigt wieder eine Anfrage, die, von einem konkreten Satz ausgehend, syntaktische Merkmale dieses Satzes erfragt.

- Zeige die NPs und ihre Position in den grammatischen Sätzen, die im zweiten Konjunkt die Konstituentenstruktur “N2_D2_A2_V2” haben:

retrieve s-id cat-desc cat-position substring where cat-desc = “np” and match(coord-desc,N2_D2_A2_V2) and wf = 1

3401 np 12 13 der student

3401 np 14 15 dem kommilitonen

3401 np 16 17 den roman

3401 np 4 5 der professor

3401 np 6 7 der sekretaerin

3401 np 8 9 den blumenstraus

3499 np 11 12 der student

...

3.4 Ausblick

DiTo bietet derzeit keinen auch nur annähernd vollständigen Überblick über die Phänomene der Koordination im Deutschen. Bisher noch gar nicht betrachtet wurden z.B. die Koordination von Adjunkten, die Koordination von Nomen (“die Männer und Frauen”), die Koordination von Präpositionen (“mit oder ohne Zucker”) und Präpositionalphrasen (“Sie spielten mit viel Einsatz, aber ohne klare Linie.”) sowie asymmetrische Koordinationen (“Er ist schnell und ein guter Springer.”). Ein weiteres noch zu berücksichtigendes Thema auf dem Gebiet der Satzkoordination sind die sogenannten “Hasensätze”:

(121) Gestern ist er gekommen und hat mit uns gegessen.

In diesen Sätzen haben die beiden Konjunkte keine symmetrische Struktur, da die Subjekt-NP im ersten, aber nicht im zweiten Konjunkt auftritt. Mögliche Erweiterungen der Datensammlung könnten diese Koordinationsphänomene behandeln.

Anhang A

Liste der möglichen Kombinationen von Tilgungen und Permutationen bei Verb-Zweit-Stellung

- Keine Tilgung: Hauptverb und 3 NPs realisiert. N V D A ¹
N V A D
D V N A
D V A N
A V N D
A V D N
6 Permutationsmöglichkeiten.
- Eine Tilgung: 3 NPs realisiert. N D A
N A D
D N A
D A N
A N D
A D N
6 Permutationsmöglichkeiten.

¹V steht für das Hauptverb, N, D und A stehen für die NPs in Nominativ, Dativ und Akkusativ.

- Eine Tilgung: Hauptverb und 2 NPs realisiert.

N V D

D V N

V N D

V D N

N V A

A V N

V N A

V A N

D V A

A V D

V D A

V A D

12 Permutationsmöglichkeiten.

- Zwei Tilgungen: 2 NPs realisiert.

N D

D N

N A

A N

D A

A D

6 Permutationsmöglichkeiten.

- Zwei Tilgungen: Hauptverb und 1 NP realisiert.

N V

V N

D V

V D

A V

V A

6 Permutationsmöglichkeiten.

- Drei Tilgungen: 1 NP realisiert.

N

D

A

3 Permutationsmöglichkeiten.

- Drei Tilgungen: Hauptverb realisiert.

V

1 Permutationsmöglichkeit.

- Vier Tilgungen: Kein Element realisiert.
0 (Der leere String) 1 Permutationsmöglichkeit.

Insgesamt gibt es für Sätze mit Verb-Zweit-Stellung 41 mögliche Kombinationen von Tilgungen und Permutationen.

Anhang B

Liste der möglichen Kombinationen von Tilgungen und Permutationen bei Verb-Letzt-Stellung

- Keine Tilgung: Hauptverb und 3 NPs realisiert. N D A V
N A D V
D N A V
D A N V
A N D V
A D N V
6 Permutationsmöglichkeiten.
- Eine Tilgung: 3 NPs realisiert. N D A
N A D
D N A
D A N
A N D
A D N
6 Permutationsmöglichkeiten.

- Eine Tilgung: Hauptverb und 2 NPs realisiert.
 N D V
 D N V
 N A V
 A N V
 D A V
 A D V
 12 Permutationsmöglichkeiten.
- Zwei Tilgungen: 2 NPs realisiert.
 N D
 D N
 N A
 A N
 D A
 A D
 6 Permutationsmöglichkeiten.
- Zwei Tilgungen: Hauptverb und 1 NP realisiert.
 N V
 D V
 A V
 6 Permutationsmöglichkeiten.
- Drei Tilgungen: 1 NP realisiert.
 N
 D
 A
 3 Permutationsmöglichkeiten.
- Drei Tilgungen: Hauptverb realisiert.
 V
 1 Permutationsmöglichkeit.
- Vier Tilgungen: Kein Element realisiert.
 0 (Der leere String) 1 Permutationsmöglichkeit.

Insgesamt gibt es für Sätze mit Verb-Letzt-Stellung 32 mögliche Kombinationen von Tilgungen und Permutationen.

Literaturverzeichnis

- István Batori:** Ein transformationelles Modell für die Koordination im Deutschen. In Batori et al. (eds.), *Syntaktische und semantische Studien zur Koordination*, Seiten 1–43. Tübingen, 1975.
- Peter Colliander:** *Das Korrelat und die obligatorische Extraposition*. Kopenhagener Beiträge zur Germanistischen Linguistik. Sonderband 2. Kopenhagen, 1983.
- Kader Diagne:** *DiTo-DMS. The DiTo Database Management System. Concepts, Implementation Issues and User Guide*. Erscheint als DFKI-Report. Saarbrücken, 1992.
- Peter Eisenberg:** *Grundriß der deutschen Grammatik*. Stuttgart, 1989.
- Ulrich Engel:** Die deutschen Satzbaupläne. In *Wirkendes Wort*, 20, Seiten 361–392, 1970.
- Ulrich Engel und Helmut Schumacher:** *Kleines Valenzwörterbuch deutscher Verben*. Tübingen, 1976
- Bernhard Engelen:** *Untersuchungen zu Satzbauplan und Wortfeld in der geschriebenen deutschen Sprache der Gegenwart*. Reihe I 3.3 Verblisten. München 1975.
- Lutz Götze:** *Valenzstrukturen deutscher Verben und Adjektive*. München, 1979.
- Gerhard Helbig:** *Wörterbuch zur Valenz und Distribution deutscher Verben*. Leipzig, 5th ed., 1980.
- Tilman N. Höhle:** On Reconstruction and Coordination. ms., 1991.
- Manfred Kohrt:** *Koordinationsreduktion und Verbstellung in einer generativen Grammatik des Deutschen*. Tübingen, 1976.

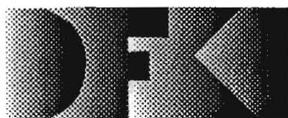
Brigitte Krenn: *DiTo. Daten-Dokumentation zu Funktionsverbgefügen.* Erscheint als DFKI-Report. Saarbrücken, 1992.

Carl Pollard und Ivan Sag: *Controlled Complements.* ms., 1990.

Ingrid Renz: *Koordination von nichtverbalen Satzkonstituenten.* IBM-IWBS Report 74. Stuttgart, 1989.

Robert R. van Oirsouw: *A Linear Approach to Coordinate Deletion,* Seiten 363 – 390. In *Linguistics*, 23, 1985.

Monika Weisgerber: *Valenz und Kongruenzbeziehungen.* Frankfurt a. M., 1983.



DFKI Publikationen

Die folgenden DFKI Veröffentlichungen sowie die aktuelle Liste von allen bisher erschienenen Publikationen können von der oben angegebenen Adresse bezogen werden.

Die Berichte werden, wenn nicht anders gekennzeichnet, kostenlos abgegeben.

DFKI Publications

The following DFKI publications or the list of all published papers so far can be ordered from the above address.

The reports are distributed free of charge except if otherwise indicated.

DFKI Research Reports

RR-91-19

Munindar P. Singh: On the Commitments and Precommitments of Limited Agents
15 pages

RR-91-20

Christoph Klauck, Ansgar Bernardi, Ralf Legleitner
FEAT-Rep: Representing Features in CAD/CAM
48 pages

RR-91-21

Klaus Netter: Clause Union and Verb Raising Phenomena in German
38 pages

RR-91-22

Andreas Dengel: Self-Adapting Structuring and Representation of Space
27 pages

RR-91-23

Michael Richter, Ansgar Bernardi, Christoph Klauck, Ralf Legleitner: Akquisition und Repräsentation von technischem Wissen für Planungsaufgaben im Bereich der Fertigungstechnik
24 Seiten

RR-91-24

Jochen Heinsohn: A Hybrid Approach for Modeling Uncertainty in Terminological Logics
22 pages

RR-91-25

Karin Harbusch, Wolfgang Finkler, Anne Schauder: Incremental Syntax Generation with Tree Adjoining Grammars
16 pages

RR-91-26

M. Bauer, S. Biundo, D. Dengler, M. Hecking, J. Koehler, G. Merziger:
Integrated Plan Generation and Recognition
- A Logic-Based Approach -
17 pages

RR-91-27

A. Bernardi, H. Boley, Ph. Hanschke, K. Hinkelmann, Ch. Klauck, O. Kühn, R. Legleitner, M. Meyer, M. M. Richter, F. Schmalhofer, G. Schmidt, W. Sommer:
ARC-TEC: Acquisition, Representation and Compilation of Technical Knowledge
18 pages

RR-91-28

Rolf Backofen, Harald Trost, Hans Uszkoreit:
Linking Typed Feature Formalisms and Terminological Knowledge Representation Languages in Natural Language Front-Ends
11 pages

RR-91-29

Hans Uszkoreit: Strategies for Adding Control Information to Declarative Grammars
17 pages

RR-91-30

Dan Flickinger, John Nerbonne:
Inheritance and Complementation: A Case Study of Easy Adjectives and Related Nouns
39 pages

RR-91-31

H.-U. Krieger, J. Nerbonne:
Feature-Based Inheritance Networks for Computational Lexicons
11 pages

RR-91-32

Rolf Backofen, Lutz Euler, Günther Görz:
Towards the Integration of Functions, Relations and
Types in an AI Programming Language
14 pages

RR-91-33

Franz Baader, Klaus Schulz:
Unification in the Union of Disjoint Equational
Theories: Combining Decision Procedures
33 pages

RR-91-34

Bernhard Nebel, Christer Bäckström:
On the Computational Complexity of Temporal
Projection and some related Problems
35 pages

RR-91-35

Winfried Graf, Wolfgang Maaß: Constraint-basierte
Verarbeitung graphischen Wissens
14 Seiten

RR-92-01

Werner Nutt: Unification in Monoidal Theories is
Solving Linear Equations over Semirings
57 pages

RR-92-02

*Andreas Dengel, Rainer Bleisinger, Rainer Hoch,
Frank Hönes, Frank Fein, Michael Malburg:*
 Π_{ODA} : The Paper Interface to ODA
53 pages

RR-92-03

Harold Boley:
Extended Logic-plus-Functional Programming
28 pages

RR-92-04

John Nerbonne: Feature-Based Lexicons:
An Example and a Comparison to DATR
15 pages

RR-92-05

*Ansgar Bernardi, Christoph Klauck,
Ralf Legleitner, Michael Schulte, Rainer Stark:*
Feature based Integration of CAD and CAPP
19 pages

RR-92-06

Achim Schupetea: Main Topics of DAI: A Review
38 pages

RR-92-07

Michael Beetz:
Decision-theoretic Transformational Planning
22 pages

RR-92-08

Gabriele Merziger: Approaches to Abductive
Reasoning - An Overview -
46 pages

RR-92-09

Winfried Graf, Markus A. Thies:
Perspektiven zur Kombination von automatischem
Animationsdesign und planbasierter Hilfe
15 Seiten

RR-92-10

M. Bauer: An Interval-based Temporal Logic in a
Multivalued Setting
17 pages

RR-92-11

Susane Biundo, Dietmar Dengler, Jana Koehler:
Deductive Planning and Plan Reuse in a Command
Language Environment
13 pages

RR-92-13

Markus A. Thies, Frank Berger:
Planbasierte graphische Hilfe in objektorientierten
Benutzungsoberflächen
13 Seiten

RR-92-14

Intelligent User Support in Graphical User
Interfaces:

1. InCome: A System to Navigate through
Interactions and Plans
Thomas Fehrle, Markus A. Thies
2. Plan-Based Graphical Help in Object-
Oriented User Interfaces
Markus A. Thies, Frank Berger

22 pages

RR-92-15

Winfried Graf: Constraint-Based Graphical Layout
of Multimodal Presentations
23 pages

RR-92-16

*Jochen Heinsohn, Daniel Kudenko, Bernhard Nebel,
Hans-Jürgen Proflich:* An Empirical Analysis of
Terminological Representation Systems
38 pages

RR-92-17

Hassan Ait-Kaci, Andreas Podelski, Gert Smolka:
A Feature-based Constraint System for Logic
Programming with Entailment
23 pages

RR-92-18

John Nerbonne: Constraint-Based Semantics
21 pages

RR-92-19

Ralf Legleitner, Ansgar Bernardi, Christoph Klauck
PIM: Planning In Manufacturing using Skeletal
Plans and Features
17 pages

RR-92-20

John Nerbonne: Representing Grammar, Meaning and Knowledge
18 pages

RR-92-21

Jörg-Peter Mohren, Jürgen Müller
Representing Spatial Relations (Part II) -The Geometrical Approach
25 pages

RR-92-22

Jörg Würtz: Unifying Cycles
24 pages

RR-92-23

Gert Smolka, Ralf Treinen:
Records for Logic Programming
38 pages

RR-92-24

Gabriele Schmidt: Knowledge Acquisition from Text in a Complex Domain
20 pages

RR-92-25

Franz Schmalhofer, Ralf Bergmann, Otto Kühn, Gabriele Schmidt: Using integrated knowledge acquisition to prepare sophisticated expert plans for their re-use in novel situations
12 pages

RR-92-26

Franz Schmalhofer, Thomas Reinartz, Bidjan Tschaitshian: Intelligent documentation as a catalyst for developing cooperative knowledge-based systems
16 pages

RR-92-27

Franz Schmalhofer, Jörg Thoben: The model-based construction of a case-oriented expert system
18 pages

RR-92-29

Zhaohur Wu, Ansgar Bernardi, Christoph Klauck: Skeletal Plans Reuse: A Restricted Conceptual Graph Classification Approach
13 pages

RR-92-33

Franz Baader
Unification Theory
22 pages

DFKI Technical Memos**TM-91-11**

Peter Wazinski: Generating Spatial Descriptions for Cross-modal References
21 pages

TM-91-12

Klaus Becker, Christoph Klauck, Johannes Schwagereit: FEAT-PATR: Eine Erweiterung des D-PATR zur Feature-Erkennung in CAD/CAM
33 Seiten

TM-91-13

Knut Hinkelmann:
Forward Logic Evaluation: Developing a Compiler from a Partially Evaluated Meta Interpreter
16 pages

TM-91-14

Rainer Bleisinger, Rainer Hoch, Andreas Dengel: ODA-based modeling for document analysis
14 pages

TM-91-15

Stefan Bussmann: Prototypical Concept Formation An Alternative Approach to Knowledge Representation
28 pages

TM-92-01

Lijuan Zhang:
Entwurf und Implementierung eines Compilers zur Transformation von Werkstückrepräsentationen
34 Seiten

TM-92-02

Achim Schupeta: Organizing Communication and Introspection in a Multi-Agent Blocksworld
32 pages

TM-92-03

Mona Singh
A Cognitive Analysis of Event Structure
21 pages

TM-92-04

Jürgen Müller, Jörg Müller, Markus Pischel, Ralf Scheidhauer:
On the Representation of Temporal Knowledge
61 pages

TM-92-05

Franz Schmalhofer, Christoph Globig, Jörg Thoben
The refitting of plans by a human expert
10 pages

TM-92-06

Otto Kühn, Franz Schmalhofer: Hierarchical skeletal plan refinement: Task- and inference structures
14 pages

DFKI Documents**D-91-16**

Jörg Thoben, Franz Schmalhofer, Thomas Reinartz: Wiederholungs-, Varianten- und Neuplanung bei der Fertigung rotationssymmetrischer Drehteile
134 Seiten

D-91-17

Andreas Becker:

Analyse der Planungsverfahren der KI im Hinblick auf ihre Eignung für die Arbeitsplanung

86 Seiten

D-91-18

Thomas Reinartz: Definition von Problemklassen im Maschinenbau als eine Begriffsbildungsaufgabe

107 Seiten

D-91-19

Peter Wazinski: Objektlokalisierung in graphischen Darstellungen

110 Seiten

D-92-01

Stefan Bussmann: Simulation Environment for Multi-Agent Worlds - Benutzeranleitung

50 Seiten

D-92-02

Wolfgang Maaß: Constraint-basierte Platzierung in multimodalen Dokumenten am Beispiel des Layout-Managers in WIP

111 Seiten

D-92-03

Wolfgang Maaß, Thomas Schiffmann, Dudung Soetopo, Winfried Graf: LAYLAB: Ein System zur automatischen Platzierung von Text-Bild-Kombinationen in multimodalen Dokumenten

41 Seiten

D-92-04

Judith Klein, Ludwig Dickmann: DiTo-Datenbank - Datendokumentation zu Verbreitung und Koordination

55 Seiten

D-92-06

Hans Werner Höper: Systematik zur Beschreibung von Werkstücken in der Terminologie der Featuresprache

392 Seiten

D-92-07

Susanne Biundo, Franz Schmalhofer (Eds.): Proceedings of the DFKI Workshop on Planning

65 pages

D-92-08

Jochen Heinsohn, Bernhard Hollunder (Eds.): DFKI Workshop on Taxonomic Reasoning Proceedings

56 pages

D-92-09

Gernod P. Laufkötter: Implementierungsmöglichkeiten der integrativen Wissensakquisitionsmethode des ARC-TEC-Projektes

86 Seiten

D-92-10

Jakob Mauss: Ein heuristisch gesteuerter Chart-Parser für attributierte Graph-Grammatiken

87 Seiten

D-92-11

Kerstin Becker: Möglichkeiten der Wissensmodellierung für technische Diagnose-Expertensysteme

92 Seiten

D-92-12

Otto Kühn, Franz Schmalhofer, Gabriele Schmidt: Integrated Knowledge Acquisition for Lathe Production Planning: a Picture Gallery (Integrierte Wissensakquisition zur Fertigungsplanung für Drehteile: eine Bildergalerie)

27 pages

D-92-13

Holger Peine: An Investigation of the Applicability of Terminological Reasoning to Application-Independent Software-Analysis

55 pages

D-92-15

DFKI Wissenschaftlich-Technischer Jahresbericht 1991

130 Seiten

D-92-16

Judith Engelkamp (Hrsg.): Verzeichnis von Softwarekomponenten für natürlichsprachliche Systeme

189 Seiten

D-92-17

Elisabeth André, Robin Cohen, Winfried Graf, Bob Kass, Cécile Paris, Wolfgang Wahlster (Eds.):

UM92: Third International Workshop on User Modeling, Proceedings

254 pages

Note: This document is available only for a nominal charge of 25 DM (or 15 US-\$).

D-92-18

Klaus Becker: Verfahren der automatisierten Diagnose technischer Systeme

109 Seiten

D-92-19

Stefan Dittrich, Rainer Hoch: Automatische, Deskriptor-basierte Unterstützung der Dokumentanalyse zur Fokussierung und Klassifizierung von Geschäftsbriefen

107 Seiten

D-92-21

Anne Schauder: Incremental Syntactic Generation of Natural Language with Tree Adjoining Grammars

57 pages

**DiTo-Datenbank - Datendokumentation zu Verbrektion und Koordination
Judith Klein und Ludwig Dickmann**

D-92-04
Document