

In: Beiträge zur Linguistik und Informationsverarbeitung Heft 19, 1969. München: Oldenbourg-Verlag, S. 7-39

## Dokumentation eines Programms zur Analyse russischer Sätze

Von J. W. JANDA, A. ROTHKEGEL, H. ZIMMERMANN, Saarbrücken

### TEIL 1

#### 1. Allgemeine Vorbemerkungen

Bei dem im folgenden beschriebenen Programm handelt es sich um den Analyseteil eines automatischen Übersetzungsverfahrens (Russisch-Englisch), das in dieser Form von P. Toma für eine IBM 360/50 konzipiert wurde. Es stellt teilweise eine Umarbeitung eines Übersetzungsprogramms dar, das für die Rechenanlage IBM 7090 geschrieben worden war. Dieses Programm wiederum leitet sich aus der GAT (Georgetown Automatic Translation) her. Sie wurde Ende der fünfziger Jahre von einer Arbeitsgruppe an der Georgetown University entwickelt, der P. Toma angehörte<sup>1</sup>.

Das hier zu beschreibende Programm wurde von der Deutschen Forschungsgemeinschaft (DFG) angekauft und der Saarbrücker Arbeitsgruppe für linguistische Datenverarbeitung übergeben, wo es bei dem Projekt einer automatischen Sprachübersetzung Russisch-Deutsch verwertet werden soll. Da dieses Verfahren bisher noch nicht beschrieben worden ist, sahen wir unsere erste Aufgabe darin, eine vollständige Dokumentation dazu herzustellen. Außer der folgenden verbalen Beschreibung wurde eine Dokumentation in Flussdiagrammform angefertigt<sup>2</sup>.

##### *1.1 Linguistische und technische Aspekte*

Das Verfahren zur maschinellen syntaktischen Analyse russischer Sätze bildet den ersten Teil eines Übersetzungsprogramms, dessen erklärtes Ziel es ist, möglichst große Mengen russischer Texte möglichst schnell in vielleicht nicht elegantes, aber verständliches Englisch (bzw. Deutsch) zu übersetzen. Es ist also ganz an der Praxis orientiert und legt wenig Wert weder auf Perfektion in der Qualität der Übersetzung noch auf die linguistische Gültigkeit der Satzbeschreibungen. Daher sind auch in der Analyse linguistische und technische Aspekte so sehr miteinander verbunden, dass es gelegentlich fast unmöglich, fast immer unzweckmäßig scheint, sie in der Beschreibung zu trennen. Das wird bereits deutlich bei der „Satzvorbereitung“. In diesem der eigentlichen Analyse vorausgehenden Teil des Programms wird der Eingabetext zunächst in Sätze unterteilt, diese ihrerseits in formale Einheiten, die „Wörter“. „Wort“ heißt dabei nicht nur jede graphische Wortform mit ihren Flexionsendungen, auch Satzzeichen gelten als Wörter, die mit den entsprechenden Wörterbuchinformationen auszustatten sind. Jedes im Text auftretende Wort wird dabei mit den Eintragungen in einem syntaktischen Wörterbuch verglichen. Technisch gesehen werden dabei meist mehrere Sätze zugleich (sog. „batches“) mit Informationen versehen. Das berührt aber nicht das Prinzip der satzweisen Analyse. Eine dieser Informationen, auf denen dann später die Analyse aufbaut, ist z. B. die Zugehörigkeit zu einer bestimmten Wortklasse. Die Einteilung in Wortklassen, die hier „Parts of Speech“ (POS) genannt werden, geht aber nun nicht nur davon aus, wie sich die betreffenden Wörter im Satz verhalten, d. h. sie folgt nicht allein der Distribution, sondern sie berücksichtigt alle möglichen Gesichtspunkte der technischen Bearbei-

tung. So gibt es auf der einen Seite die Parts of Speech „Substantiv“, „Adverb“, „finites Verb“, die als syntaktische Klasse anzusetzen linguistisch sicher gerechtfertigt ist, auf der anderen Seite gibt es aber auch die „Parts of Speech“: „Paragraph“ „zu übergehendes Wort“, „CTO“<sup>2a</sup> (d. h. ein einzelnes Wort) usw. Dieses praktisch durchaus zu rechtfertigende Vorgehen schadet natürlich der linguistischen Systematik und führt z. T. auch zu komplizierten Bearbeitungsverfahren während der Analyse.

Es ist auch niemals der Versuch gemacht worden, die linguistische Basis zu definieren.

Was die Analyse prinzipiell zu leisten hat, lässt sich durch den Begriff „Reduktion“ andeuten. Aus einer im Text aufgefundenen formalen Einheit, einem „Wort“ im angegebenen Sinn, ist nicht unmittelbar abzulesen, welche Funktion sie im Satz erfüllt. Das aber muss für die Übersetzung bekannt sein, denn in der Zielsprache muss diese Funktion ja mit eventuell ganz anderen syntaktischen Mitteln wiedergegeben werden. Die Funktionsbestimmung ist nur teilweise über ein Wörterbuch lösbar, denn welche Funktion ein „Wort“ im konkreten Satz erfüllt, hängt größtenteils von seiner Umgebung ab. Das Wort „Vater“ z. B. kann Subjekt, Akkusativobjekt, Dativobjekt, Prädikatsnomen u. a. sein. Die Analyse soll diese verschiedenen möglichen Funktionen auf eine, nämlich die im Satz gegebene, reduzieren. Diese Reduktion erfolgt auf verschiedenen Ebenen; ein „Wort“ kann in verschiedener Weise als Element einer Wortgruppe fungieren, diese wiederum verschieden als Teil eines Satzgliedes usw. Aber diese aufeinander folgenden Ebenen oder, wie man auch sagen könnte, „Reduktionsstufen“ werden in den einzelnen Bearbeitungsschritten nie scharf getrennt, etwa so, dass man die Stufen

Wortform-Wortklasse-Gruppe- Satzglied- Satzteil- Satz

unterschiede und jeder einen Programmschritt, etwa einen „Durchlauf“ (Pass) durch den Satz, zuteilt. Zwar gibt es fünf aufeinander folgende Durchläufe (mit wechselnder Bearbeitungsrichtung); aber sie entsprechen nicht den Reduktionsstufen. Vielmehr wird in jedem Durchlauf möglichst viel zu bestimmen versucht. Hat man etwa in Pass 1 festgestellt, dass ein bestimmtes Substantiv (SUB) und ein bestimmtes Adjektiv (ADJ) zusammengehören und sie entsprechend zugeordnet, so wird nicht gleich zum nächsten Wort gegangen; vielmehr wird untersucht, ob das betreffende SUB vielleicht Subjekt des Satzes ist usw. Prinzipiell lässt sich gegen ein solches Vorgehen natürlich nichts einwenden; es entspringt gewissen praktischen Notwendigkeiten (z. B. Einsparung von Rechenzeit). Aber es führt dazu, dass bereits Reduktionen erfolgen, ohne dass alle dafür relevanten Informationen ermittelt sind. Und da es bisher nicht möglich ist, falsche Reduktionen aus einem früheren Umlauf in einem späteren zu korrigieren, kommt es häufig zu falschen Analysen allein deshalb, weil bestimmte Entscheidungen über die Funktion einer bestimmten Einheit zu früh gefällt worden sind. Beispielsweise sei im 2. Durchlauf (= Pass 1), der von rechts nach links geht, das 6. Wort von rechts ein mögliches Subjekt; aus der Analyse der Wörter 1 bis 5 (von rechts) hat sich nichts ergeben, was dagegen spräche, also wird Wort 6 zum Subjekt erklärt, obwohl die Untersuchung von Wort 7 vielleicht gezeigt hätte, dass das im vorliegenden Satz nicht der Fall ist. Hier erweist sich die technische Vermischung der einzelnen Reduktionsstufen als Mangel; in anderen Fällen ist sie unvermeidlich, wenn falsche Reduktionen vermieden werden sollen. Es erscheint aus den erwähnten Gründen wenig sinnvoll, eine rein linguistische Beschreibung aus dem Programm zu extrahieren; es ist aber sicher zweckmäßig, die wichtigsten linguistischen Einheiten, die durch die Analyse bestimmt werden können, hier in einer Liste zusammenzustellen. Wir bezeichnen sie mit traditionellen Termini und erläutern sie durch deutsche Beispiele:

- |   |  |
|---|--|
| a) Subjekt/Objekt/Prädikativum                        |  |
| einfache Nominalgruppe (NOG)                          | der (junge) Mann                                   |
| angereihte Nominalgruppe (NOG)                        | Männer und Frauen                                  |
| erweiterte Nominalgruppe (NOG)<br>(mit Genitivattr.)  | der Freund des Mädchens                            |
| verkettete Nominalgruppe (NOG)<br>Präpositionalgruppe | die den Mann sehende Frau<br>mit seinem Freund     |
| b) Prädikat   |  |
| Verbalgruppe  | sieht, kann sehen                                  |
| c) Satztypen  |  |
| Hauptsatz   | er kommt   |
| Nebensatz   | weil er kommt                                      |
| Fragesatz   | kommt er?  |
| angereihter Nebensatz                                 | (weil er kommt) und sie geht                       |
| Gerundivsatz  | die Augen grässlich verdrehend, ..                 |
| Präpositionalersatz                                   | (Er blieb,) trotz vieler Kontaktversuche, (einsam) |
| Einbettung (Relativsatz)                              | ..., der dort kommt ...                            |
| d) Vergleich  | (stärker) als Achill                               |
| e) Anreihungen allg.                                  | Peter, Irma und Pauline                            |

## 1.2 Termini und Abkürzungen

Zum Verständnis der folgenden Beschreibung ist es erforderlich, die wichtigsten, stets wiederkehrenden Termini zu kennen. In Klammern geben wir gegebenenfalls eine deutsche Übersetzung.

Bei Zuordnungen zweier Wörter werden stets ein „regierendes“, der Governor (Leitwort, auch Principle Word), und ein „regiertes“, der Satisfier (Sättiger, auch Supplement), unterschieden. Bei der Objektzuordnung ist z. B. das Verb Governor, das entsprechende Objekt Satisfier, bei Präpositionalgruppen regiert die Präposition (= Governor) das entsprechende Substantiv (= Satisfier).

Je nach Position des Satisfier unterscheidet man Right Agreement (Zuordnung rechts, d. h. der Satisfier steht rechts vom Governor) und Left Agreement (Zuordnung links, d. h. der Satisfier steht links vom Governor). Allerdings werden diese Termini meist nur bei Zuordnungen innerhalb nominaler Gruppen (SUB-AD J oder AD J-SUB) sowie bei der Zuordnung Adverb-Verb gebraucht. Diese Begriffe haben neben der grammatischen eine maschinentechnische Bedeutung. Vor allem soll dadurch das Aufsuchen von Informationen erleichtert werden. Besteht das Subjekt z. B. aus einer vielfach erweiterten Nominalgruppe, so braucht der Hinweis „Subjekt“ nur einmal, nämlich beim Governor, aufgenommen zu werden. Zusätzlich enthält der Governor Angaben, welche Einheiten er regiert.

Die Analyse erfolgt, wie schon mehrfach erwähnt, Wort für Wort. Das Wort, das gerade behandelt wird und von dem es abhängt, welcher Programmzweig angesprochen wird, nennen wir Ausgangswort (AW).

*Abkürzungen:*

AD J	Adjektiv
ADV	Adverb
AKK	Akkusativ
DAT	Dativ
GEN	Genitiv
IBPQ	Identification Basic Paradigmatic Cue
INF	Infinitiv
NOG	nominale Gruppe
NOM	Nominativ
POS	Part of Speech
PTZ	Partizip
VRB	Verb
ZO	(nominale) Zuordnung

## **2. Der Analysespeicher**

Die gesamten Informationen zu jedem Wort eines Satzes sind in vierundsechzig Bytes zu je acht Bits gespeichert. Der laufende Text wurde deshalb im Ausdruck des Analysespeichers nicht mehr aufgeführt. Je Byte werden zwei Hexadezimalzahlen geschrieben. Es gibt zwei Arten von Eintragungen: Einmal wird das Byte als Ganzes interpretiert, es enthält dann eine Wortbezugsnummer (= Wortnummer eines Textwortes im Satz), verschiedentlich sind die Bits eines Bytes auch einzeln zu interpretieren. In letzterem Fall wird das jeweilige Bit bei der Beschreibung angegeben.

### *2.1 Übersicht*

In Byte 0 ist die laufende Nummer des Wortes im Satz aufgeführt. Der Teil des Analysespeichers, der die Bytes 1 bis 12 umfasst, enthält grammatische Informationen, die größtenteils bei Beginn der Satzanalyse bereits vorliegen und im allgemeinen aus dem Wörterbuch übernommen sind.

In den Bytes 13 bis 15 wird die Adresse angegeben, unter der die Syntheschiffrierung zu finden ist.

Die aus der Satzanalyse resultierenden Ergebnisinformationen werden in den Bytes 16 bis 55 verzeichnet. Jedoch werden bereits hier teilweise zur Synthese gehörende Eintragungen vorgenommen (vgl. unten bei der ausführlichen Beschreibung der einzelnen Bytes). Die mit Byte 56 beginnenden Angaben betreffen nur noch die Synthese. In den Bytes 61 bis 63 werden während der Synthese die Adressen aufgeführt, an denen die Übersetzungen der betreffenden Wörter zu finden sind.

## 2.2 Ausführliche Beschreibung

- Byte 0: Nummer des Wortes im Satz
- Byte 1: Angabe der Wortklasse (Part of Speech, zum einzelnen vgl. S. 23ff.)
- Byte 2: *Nomen im Singular*  
80 = Nominativ  
40 = Genitiv  
20 = Dativ  
10 = Akkusativ  
08 = Instrumental  
04 = Präpositional
- Finites Verb Singular.*  
80 = 1. Person  
40 = 2. Person  
20 = 3. Person
- Plural:*  
10 = 1. Person  
08 = 2. Person  
04 = 3. Person
- Byte 3: *Nomen im Plural*  
*Kasusangaben wie in Byte 2*  
*Finites Verb, Gerundium, Partizip*  
80 = Präsens  
40 = Präteritum  
20 = Futur  
10 = Imperativ
- Byte 4: *Genus und Numerus*  
80 = maskulinum  
40 = femininum  
20 = neutrum  
02 = Plural  
01 = Singular
- Byte 5: *direkte Rektion*  
40 = Genitiv  
20 = Dativ  
10 = Akkusativ  
08 = Instrumental  
04 = Präpositional
- Byte 6: *Verb*  
80 = Infinitivreaktion des Leitworts  
40 = Infinitiv kann direktes Objekt ersetzen  
20 = Wort ist reflexiv

10 = Leitwort kann durch Adverb befriedigt werden  
 08 = ausschließlich impersonal gebrauchtes Verb  
 04 = nur in der 3. Person Singular Neutrum Präteritum gebrauchtes Verb (impersonale Form)  
 02 = Kurzform Neutrum des Partizips kann nur Adverb sein  
 01 = Kurzform Neutrum des Partizips nur, wenn finites Verb im Satz vorhanden ist

*Adjektiv*

10 = Adjektiv ist substantiviert und regiert Nominalphrase im Dativ  
 08 = Adjektiv ist substantiviert und regiert Nominalphrase im Instrumental  
 04 = Adjektiv wird relativisch gebraucht  
 02 = Adjektiv steht vor einem den bestimmten Artikel verlangenden Substantiv  
 01 = prädikative Kurzform Neutrum Singular ist stets Prädikat

*Pronomen*

08 = Personalpronomen

Byte 7:

*Verb*

80 = unvollendeter Aspekt  
 40 = vollendeter Aspekt  
 20 = Akkusativ wird durch Präpositionalphrase mit CEREZ, SKVOZ6, POD, NAD, V ersetzt  
 10 = Kurzform des Adjektivs kann Prädikat sein  
 04 = Kategorie der Belebtheit oder Unbelebtheit beim Akkusativobjekt  
 02 = bearbeitetes Wort ist „Hauptwort“  
 01 = bearbeitetes Wort ist Supplement in `Limited Semantics' (zu den `Limited Semantics' gehören Wörter, die in idiomatischen Wendungen oder in Fachsprachen von der Grundbedeutung abweichende Begriffe bilden)

Byte 8:

*Angabe der Wortklasse nach IBPQ (= alter Wortklassenkodex, vgl. S. 25)  
 Kennzeichnung des Nebensatzanfangs bei angereihten Nebensätzen ('FF')*

Byte 9 und

*Aufzählung der gebräuchlichsten Präpositionen*

Byte 10:

	Byte 9	Byte 10
80	DL4	U
40	K, KO	V, VO
20	OT	ZA
10	NA	O, OB, PRO, KASATEL6NO, OTNOSITEL6NO, NASCET
08	PO	PRI
04	POD	IZ, IZO
02	PROTIV	DO
01	S, SO	CEREZ

Byte 11:

*Kategorie der Belebtheit und Unbelebtheit*

80 = Wörter für Menschen, die bei Wortverbindungen mit Wort für unbelebte Wesen Gruppe von Menschen bedeuten

40 = Wörter für belebte Wesen  
20 = Wörter für unbelebte Wesen in zu Byte 80 analoger Wortverbindung bedeuten Gruppe von Menschen

*Konjunktionen*

08 = CTOBY  
04 = CTOBY, KOGDA  
02 = CTO  
01 = CTOBY, KOGDA, GDE, OTKUDA, POCEMU, ZACEM, CTO, KTO, KAKO1

- Byte 12: *Wortklassen werden Zu Grundklassen zusammengefasst*  
80 = Substantiv  
40 = Adjektiv in Kurzform  
20 = Partizip in Kurzform  
10 = Verbalwort  
08 = Adverb  
04 = Bindewort  
02 = Adverb oder Kurzform des Adjektivs Neutrum Singular. Bei attributiven Langformen der Adjektive und Partizipien wird zu 40 beziehungsweise 20 noch 01 hinzugefügt.
- Byte 13-15: vgl. Übersicht 2. 1
- Byte 16: *Zuordnung links*  
Die links vor dem regierenden Wort diesem zuordnungsfähigen Wörter erhalten nach erfolgter Zuordnung die Wortnummer des regierenden Wortes ('Left Agreement' = Zuordnung links)
- Byte 17: *Zuordnung rechts*  
Analog zu Byte 16 wird das regierende Wort und die rechts nach diesem stehenden zuordnungsfähigen Wörter zusammengefasst ('Right Agreement' = Zuordnung rechts)
- Byte 18 und  
Byte 19: Bei Objektzuordnung erhält das regierende Wort die Wortnummer des ersten bzw. des zweiten Objekts
- Byte 20: Wortnummer des Genitivattributs bei regierendem Wort wird eingetragen
- Byte 21: Infinitivreaktion; enthält die Wortnummer des Infinitivs bei regierendem Wort
- Byte 22 und  
Byte 23: enthalten die Wortnummer der Nebensatzeinleitenden Konjunktionen (regiertes Wort) CTO beziehungsweise CTOBY usw. bei regierendem Wort
- Byte 24 und  
Byte 25: enthalten die Wortnummer des ersten beziehungsweise des zweiten Präpositionalobjekts bei regierendem Wort
- Byte 26 und: enthalten beim 'Governor' die Wortnummer des regierten Wortes

- Byte 27: (entsprechend Byte 16 und 17)
- Byte 28 und  
Byte 29: enthalten die Wortnummer des regierenden Wortes für Objektzuordnung bei regiertem Wort (vgl. Byte 18 und 19)
- Byte 30: enthält die Wortnummer des regierenden Wortes bei Genitivattribut (vgl. Byte 20)
- Byte 31: enthält die Wortnummer des regierenden Wortes bei Infinitiv (vgl. Byte 21)
- Byte 32 und  
Byte 33: enthalten entsprechend Byte 22 und 23 Zusammenfassung von regierendem Wort und Konjunktion. Byte 32 enthält außerdem den Hinweis, ob auf Grund von Kasusmehrdeutigkeit mehrfache Subjekte möglich sind
- Byte 34: bei eingeschobenen Sätzen enthält jedes Wort des Satzes die Eintragung 'FF'
- Byte 35: bei Präpositionalobjekt Hinweis auf Präposition
- Byte 36: *Subjektanzeiger*  
80 = 'sicheres' Subjekt  
40 = 'mehrfaches' Subjekt  
20 = 'mögliches' Subjekt
- Substantiv*  
10 = Substantiv nach 3T0 oder KAK  
04 = Substantiv nach CEM (ALS)  
02 = Substantiv ist Apposition
- Byte 37: *Prädikatanzeiger*  
80 = 'sicheres' Prädikat  
40 = 'mögliches' Prädikat
- Byte 38: *Angaben zum regierten Wort*  
80 = direkte und präpositionale Rektion  
40 = Genitivattribut  
20 = abhängiger Infinitiv  
10 = CTO  
08 = Zuordnung nach links  
04 = Zuordnung nach rechts  
01 = andere regierte Wörter  
Außerdem enthält Byte 38 Angaben zur Funktion des Gedankenstrichs im Satz
- Byte 39: enthält analog zu Byte 38 Eintragungen zum regierenden Wort Ausnahme (keine Angaben zum Gedankenstrich)  
01 = Funktion des Worts wurde bestimmt
- Byte 40: *Aufzählungen*  
80 = Kommata

40 = Adverbien vom Typ MNOGO und prädikative Kurzformen des Adjektivs  
20 = Infinitive  
10 = CTO  
08 = Substantive  
04 = Hilfsverben  
*Komma*  
02 = Komma einer Präpositionalphrase  
01 = Komma einer Partizipial- oder Gerundphrase

Byte 41: *Kasusmehrdeutigkeit bei Substantiv*  
04 = Genitivattribut könnte auch der Wortform nach im Akkusativ gebraucht sein  
02 = Genitivattribut kann auch im Nominativ gebraucht sein  
01 = Substantiv hat zusätzlich Nominativbit

Byte 42: *mögliche Kasus des Nomens vor der Kasusreduktion*  
80 = Nominativ  
40 = Genitiv  
20 = Dativ  
10 = Akkusativ  
08 = Instrumental  
04 = Präpositional

*finites Verb*

Angaben zur englischen und deutschen Synthese (Geschlecht der Person in der Zielsprache) deutsch  
08 = maskulinum  
04 = femininum  
02 = neutrum englisch  
80 = maskulinum  
40 = femininum  
20 = neutrum

Byte 43: *Nomen*  
Eintragungen zu Byte 42 analog nach erfolgter Kasusreduktion

Byte 44: *Bestimmung des Genus und Numerus*  
80 = maskulinum  
40 = femininum  
20 = neutrum  
02 = Plural  
01 = Singular

*Angaben zur Synthese*

10 = eventuell einzufügendes Personalpronomen  
08 = bestimmter Artikel einzufügen  
04 = unbestimmter Artikel einzufügen

- Byte 45: die in Byte 5 aufgeführten, jedoch noch nicht befriedigten möglichen Rektionen werden aufgeführt
- Byte 45 und Byte 46: von den in Byte 5 und 6 aufgeführten syntaktischen Fügungen werden nur noch die unerfüllten Forderungen verzeichnet. Weiterhin wird in Byte 45 an der Satzgrenze bei Konjunktion oder Satzzeichen die Wortnummer des ersten Prädikats angegeben
- Byte 47 und Byte 48: weist durch Angabe der Wortnummer hin auf das rechts (47) beziehungsweise links (48) stehende Wort der gleichen Wortklasse innerhalb einer Aufzählung
- Byte 49: *Finites Verb*  
bei Verbalgruppe erhält finites Verb die Wortnummer des rechts stehenden Infinitivs
- Byte 50: *Zusatzinformationen*  
80 = Konditionalsatz  
40 = Fragesatz  
20 = Negation  
08 = bearbeitetes Wort kann nicht als regiertes Wort fungieren  
04 = bearbeitetes Wort kann nicht als Genitivattribut fungieren  
01 = Imperativ der ersten oder zweiten Person
- Byte 51: mögliche Grenzanzeiger für Satzeinheiten werden durch 'FF' gekennzeichnet
- Byte 52: Bezug zwischen Relativwort und Bezugswort Zusammenfassung der Prädikate (vgl. Byte 47)
- Byte 53: *Eintragungen zur Synthese*  
80 = in der Zielsprache wird Wort übergangen  
40 = in der Zielsprache wird Partizip umgestellt
- Angaben zur Analyse*  
40 = Komma vor CEM  
20 = Komma zwischen zwei Wörtern der gleichen Wortklasse
- Byte 54: bei Schachtelung von Nebensätzen Nummer des ersten Wortes des fortgesetzten Subsatzes
- Byte 55: *Angaben zum Subjekt*  
Ist nur ein Subjekt vorhanden, wird dessen Wortnummer eingetragen; gibt es mehrere Subjekte, die keine Verben sind, so erfolgt Eintrag `FF'
- Byte 56 und Byte 57: enthält für Wortumstellung in der Zielsprache Nummer des ersten beziehungsweise des letzten Wortes
- Byte 58: Übersetzung der Präpositionen

Byte 59: Die für den Druck verbindliche Wortfolge wird festgehalten. Dabei erhält das Wort im Satz die Nummer des zuerst zu druckenden Wortes usw.

Byte 60: Länge des Ausdrucks (in Bytes) des vorliegenden Satzes

Byte 61 und

Byte 63: Vergleiche Übersicht 2. 1

### **3. Beschreibung der Analyseschritte**

#### *Pass 0: Erste Reduktion von Kasusmehrdeutigkeiten*

Im Durchlauf 0 wird nach Möglichkeit eine erste Kasus- und Genusreduktion bei Substantiven durchgeführt. Wenn ein Substantiv und ein Adjektiv vor diesem Substantiv in bestimmten Kasus (oder nur in einem Kasus) übereinstimmen, wird eine evtl. darüber hinausgehende Kasusmehrdeutigkeit des SUB auf die diesem SUB und dem AD J gemeinsamen Kasusmöglichkeiten reduziert. Das Ergebnis wird bei dem entsprechenden SUB (Byte 43) vermerkt. Dasselbe geschieht für die Genusangaben (Ergebnis in Byte 44).

Dieser Programmschritt bearbeitet den Satz vom ersten zum letzten Wort hin, d. h. von links nach rechts. Zunächst wird ein Substantiv gesucht, dem ein Adjektiv oder Partizip vorausgeht. Bevor eine mögliche Übereinstimmung im Kasus und Genus abgefragt wird, werden solche Kombinationen ausgeschlossen, in denen das Substantiv im Singular, das Adjektiv aber im Plural auftritt. Hier ist zu diesem Zeitpunkt keine Reduktion möglich, da sich das AD J auf ein anderes SUB beziehen kann.

Wenn keine Übereinstimmung vorhanden ist, wird das nächste SUB im Satz gesucht und das Programm dafür wiederholt. Ist die Bearbeitung des Satzes beendet, erfolgt der Rücksprung in das Hauptprogramm. Dieser Analyseschritt wurde eingefügt, nachdem es sich gezeigt hatte, dass bei der in Pass 1 erfolgenden Objekt-Prädikat-Zuordnung unter gewissen Voraussetzungen Fehler auftraten: Pass 1 bearbeitet den Satz vom letzten zum ersten Wort hin. Es ist möglich, dass dabei ein Verb auftritt, bei dem rechts noch nicht alle Rektionen (= Objektmöglichkeiten) erfüllt sind. Wenn (links) vor ihm ein Substantiv auftritt, dürfen nicht die (häufig mehrdeutigen) Deklinationsangaben des SUB die Grundlage des Versuchs einer Objekt-Zuordnung bilden, sondern die übereinstimmenden Angaben einer evtl. vorhandenen nominalen Gruppe, die zu diesem Zeitpunkt in Pass 1 jedoch noch nicht erkannt ist. Eine ähnliche Priorität hat auch die Zuordnung zwischen Präpositionen und Substantiven. Vor einer Objekt-Zuordnung müsste geklärt sein, ob das Substantiv bereits von einer Präposition regiert wird, da es in diesem Fall kein direktes Objekt sein kann, auch wenn es dem Kasus nach möglich wäre. Ein entsprechend geplanter Programmschritt, der in Pass 0 einzufügen gewesen wäre, war zum Zeitpunkt der Programmübergabe noch nicht ausgefertigt<sup>3</sup>.

#### *Pass 1: Nominale und verbale Gruppierung Überblick*

In Abhängigkeit von den Wortklassen wird ein Satz vom letzten zum ersten Wort hin, also von rechts nach links fortschreitend, untersucht<sup>4</sup>. Dabei werden bereits alle erkennbaren Strukturen (nominale Gruppen - evtl. Subjekte, Objekte, Präpositionalgefüge, Attribute - und verbale Gefüge - Prädikate) festgehalten. Die Analyse basiert dabei auf Stellungs- und Zuordnungsregeln und interpretiert daneben die zu den Wortklassen gefügten Informationen des maschinellen Wörterbuchs, z. B. Flexionsangaben, Valenzen (= Hinweise für den Kasus, in dem die möglichen direkten Objekte stehen). Die Ergebnisse werden einmal im sog. Analyse-Speicher (Analysis Area) bei jedem Wort in entsprechend definierten Feldern aufbewahrt (etwa bei Attributen der Hinweis auf das Hauptwort einer Nominalgruppe, bei Objekten der Verweis auf das Prädikat). Daneben werden während des Durchlaufs beim Auftreten bestimmter Informationen Hinweisschalter (Switches) gesetzt, die über das bereits festgestellte und für die weitere Analyse unmittelbar wichtige Inventar des Satzes Auskunft geben. Zu diesen Schaltern gehört im Allgemeinen ein Zeiger (Pointer), der auf das die entsprechende Information enthaltende Wort oder Gefüge verweist. Die Analyse benutzt also neben den Wörterbuchangaben diese bereits erarbeiteten Informationen, d. h. sie berücksichtigt das Inventar, das rechts nach dem Wort vorhanden ist, das gerade bearbeitet wird (Ausgangswort). Nur in Ausnahmefällen werden in Pass 1 Informationen links des Ausgangswortes abgefragt, etwa zur Feststellung einer Negation (NE). Die Schalter haben zusätzlich die Funktion, an die einem Wort noch fehlenden Ergänzungen zu erinnern, die dann bei einem (links) vor dem Wort auftretenden Sättiger (Satisfier) oder Leitwort (Governor) rückwirkend eingetragen werden können. Es ist auch möglich, dass beim weiteren Fortschreiten „Erinnerungen“ (d. h. Schalter) an ein bestimmtes Inventar „gelöscht“ werden, falls feststeht, dass es für die übrige Satzanalyse nicht mehr von Bedeutung ist.

Liste der verwendeten Schalter (Switches, endend auf SW) und Zeiger (Pointer, endend auf P)

Name	Funktion
BYSW	By-Partikel vorhanden (Konditionalsatz)
BYSWP	Wortnummer von BY
CEM SW	CEM vorhanden (Vergleich)
CEMP	Wortnummer von CEM (ALS)
COMMASW	Komma rechts
COMMAP	Wortnummer des Kommas (die Satzzeichen werden als Wörter betrachtet)
CTO SW	CTO vorhanden CTOPOINT Wortnummer von CTO
ILI SW	Konjunktion (I, ILI)
ILIP	Wortnummer der Konjunktion
INFSW	Infinitiv rechts vorhanden
INFP	Wortnummer des Infinitivs
LAGUSW	Nominale Zuordnung (Left Agreement) durchgeführt
LAGUSWP	Wortnummer des Bezugswortes
LAPRINWP	Hauptwort bei `Zuordnung links'
LAGUNUM	0, wenn Kasusübereinstimmung AD J + SUB im Plural, sonst 1
LI SW	LI-Partikel vorhanden (Fragesatz)
LISWP	Wortnummer von LI
MNOGOSW	Adverb vom Typ MNOGO vorhanden
MNOGOP	Wortnummer des Adverbs vom Typ MNOGO
NGOVR	Substantiv von rechts regiert (z. B. als Objekt vom Verb)
NGOVRP	Wortnummer des SUB, das von rechts regiert wird

NOUNSW	Substantiv rechts vorhanden (auch bei Personalpronomen, das nicht im Nominativ steht)
NSWP	Wortnummer des SUB rechts
N SWORG	1, wenn Einsprung in UP erfolgt auf Grund NOUN SW, sonst 0
POTLGSW	Potential Left Government: Noch unerfüllte Valenzen des Prädikats
PLGOVP	Wortnummer des Prädikats mit noch unerfüllten Valenzen
RAGORG	1, wenn Einsprung in UP erfolgt auf Grund NOUNSW, sonst 0
RAGRSW	Right Agreement möglich: Freies AD J/PTZ rechts
RAGRSWP	Wortnummer des Wortes, bei dem `Zuordnung rechts' möglich ist
RLAGUSW	Vorletztes 'Left Agreement': Reserveanzeige für erfolgte Zuordnung
RLAGSWP	Wortnummer des Hauptwortes bei vorletzter nominaler Zuordnung
RNOUNSW	Substantiv vorhanden (Reserve: vorletztes SUB)
RNSWP	Wortnummer des vorletzten SUB rechts
RN SWORG	1, wenn Einsprung in UP erfolgt auf Grund RNOUNSW, sonst 0
SHFAPSW	Short-Form-ADJ / PTZ: Kurzform-Adjektiv / Partizip vorhanden
SHFAPSWP	Wortnummer des Kurzform-AD J/PTZ rechts
UK2SW	1, wenn MNOGO 2. Satisfier ist, sonst 0
VERBFSW	rechts Verb vorhanden (auch Infinitiv oder Gerundium)
VERBFP	Wortnummer des Verbs rechts

Im Durchlauf 1 verzweigt sich das Analyseprogramm in 27 verschiedene Programmteile, in denen 41 Wortklassen (Parts of Speech, POS) direkt angesprochen und bearbeitet werden. Die restlichen 17 POS werden in Pass 1 nicht bearbeitet. Bereits bei der Satzvorbereitung sind die insgesamt 58 POS automatisch aus den 100 „alten“ Wortklassenkodierungen (Identification Basic Paradigmatic Cue, IBPQ) des Wörterbuchs<sup>5</sup> gewonnen worden. Dabei wurden beispielsweise die im IBPQ als einzelne „Wortklassen“ aufgeführten Präpositionen (Codes AO - AF und BO - B3) zur Wortklasse „Präposition“ (POS 50) zusammengefasst. Die im IBPQ aufgeführten 24 Wörter mit verschiedenen Bedeutungen (Codes CO -D7) treten in den POS nicht gesondert auf. Desgleichen fehlen vier weitere IBPQ (Paragraph Mark 97, Title Mark 98, LIBO 6C, VOT 6D), die alle als POS die Angaben DC erhalten (Disregarded Word). Es sei angemerkt, dass die Analyse zwar auf den Parts of Speech basiert, die IBPQ aber immer zusätzlich herangezogen werden können. Die Einteilung in POS scheint in erster Linie aus Gründen der leichteren Verarbeitung erfolgt zu sein. Wir geben im folgenden eine Übersicht mit den Verzweigungen in Pass 1 und den beteiligten Wortklassen POS und IBPQ (GENRET = GENeral RETurn = keine Bearbeitung in Pass 1).

Programmzweig	POS	IBPQ	Bedeutung engl. (Dt.) (Name)
FV100	04	41	Finite Verb (Vollverb)
		42	Finite Verb reflexive
SHFAP10	08	24	Short-Form-Adjective (Kurzform) außer Neutr. Sing.
		82	Short-Form Participle
		37	Short-Form Mask. 3rd P.Pl.
NETFPR	OC	66	NET (NEIN)
NOUN100	10	10	Pure Noun (Substantiv allg.)
		14	Proper Noun (Eigename)

	18	12	Profession or Title (Berufsbezeichnung)
ABBREV10	1C	13	Abbreviation (Abkürzung) anschl. NOUN 100
AD JPAR10	20	20	Pure Adjective (Adjektiv allg.)
		28	Superlative Adjective (Adj. superl.)
	24	21	Noun/Adjective-Ambiguity ( <i>Mehrdeutigkeit</i> Adj/ Sub)
	28	22	Pronoun Adjective (Pronominaladjektiv)
	2C	25	Literal Ordinal Numbers (Ordinalzahlen: Ziffern)
		26	Written Ordinal Numbers (geschrieb. Ord.)
	80	83	Past Passive Participle (Partizip Perfekt Passiv)
	84	84	Present Passive Participle (Partizip Präs. Passiv)
	88	85	Past active Participle (PTZ Perf. Aktiv)
		86	Past active Participle (refl)
	8C	87	Present Active Participle (PTZ Präsens Aktiv)
		88	Present Active Participle (refl)
ADVERB10	30	30	Adverb (außer MNOGO-, MALO-Type)
		32	Old Adverb Code
		35	TOL6KO (NUR)
		6B	E5E (NOCH)
ADV3110	34	31	MNOGO-, MALO-Type Adverbs (VIEL, WENIG)
SHFNEUTS	38	23	Short-Form Neuter Singular Adjective (Kurzform Neutr. S.)
		81	Short-Form Neuter Singular Participle
(GENRET)	3C		JE (DENN)
AUX100	40	45	Finite Auxiliary (Finites Hilfsverb)
	44	46	Infinite Auxiliary
INF100	48	43	Infinitive (Infinitiv)
		44	Infinitive reflexive
GERUND10	4C	80	Gerund (Gerundium)
		89	Gerund reflexive
PREP10	50	51	Old Preposition Code (Präposition)
		52	Old Preposition Code
			dazu: AO bis B3: 20 einzelne Präpositionen
(GENRET)	54	65	NE (NICHT) wird in Pass 1 von anderen Wortklassen aus bearbeitet
(GENRET)	58	47	EST6 (Homographie ESSEN/IST)
ANOXOT10	5C	63	Conjunctions (A, NO, XOT4) (Konj.)

		64	No-Agreement-Conjunctions (TOGDA, TAK, T.P., T.K., T.E., T.D.)
COMMA 10	60	09	Comma
ILII10	64	62	1 (UND) ILI (ODER) Connective Conjunctions
(GENRET)	60	63	Other Conjunctions (z. B. DA)
PERSPRSI	70	72	Personal Pronouns (only Subject) Personalpronomen im Nominativ
PRONNSUB	74	73	Pers. Pronouns (never Subject) Personalpronomen nicht Nominativ
EGOEIIX	78	75 7B 7C 7D	EE, EGO, IX (Possessivpronomen/Pers. Pron.) EE (SEIN) EGO (IHR) IX (IHR)
KOTOR10	7C	76	KOTORY1 (+KAKO1) (Relativwort)
CEM100	90	74	CEM (ALS) mit Komparativ
COMPAR10	94	27 29 2B 2A 38	Irregular Comparative Adjective Comparative BOLEE MENE Comp. Neutr. Ambiguity: (Comparative/ Neutr. s. Nom/Acc) Ambiguity Comp./Participle
BY100	98		Konditional-Partikel
LI100	9C		Frage-Partikel
NUMBERIO	AO	14 15 16 17 18	Literal Cardinal Number 1 Written Cardinal Numbers Literal Cardinal Numbers ending in 1, but not 11 Literal Cardinal Numbers ending in 11 All other Cardinal Numbers
(GENRET)	A4	17	Literal Cardinal Numbers ending in 11 (entfällt, vgl. AO)
(GENRET)	A8	60 67 6A	DAJE (SOGAR) Particle DA (JA)

(GENRET)	AC	68	Interjection
CTO 100	BO	71	CTO (Pronouns - Subject or not Subject)
SWCLBIO	B4	59	CTOBY, KOGDA
	B8	33	KAK
	BC		andere Konjunktionen, noch unbesetzt
(GENRET)	CO	03	Dash (Gedankenstrich)
(GENRET)	C4	07	Colon (Doppelpunkt)
(GENRET)	C8	93	Quotation Mark (Anführungszeichen)
(GENRET)	CC	05	Period (Punkt)
		06	Semicolon (Strichpunkt)
		08	Exclamation Mark (Ausrufezeichen)
		92	Question Mark (Fragezeichen)
(GENRET)	DO	01	Open Bracket (Geöffnete Klammer, eckig)
		90	Open Parenthesis (Öffnende Klammer, rund)
(GENRET)	D4	02	Closed Bracket (Schließende Klammer, eckig)
		91	Closed Parenthesis (Schließende Klammer, rund)
(GENRET)	D8	04	Dots (Punkte)
(GENRET)	DC		Disregarded Words
(GENRET)	EO		Words not found in Dictionary
(GENRET)	E4	60	DA JE: Adverb/Conjunction Particle Ambiguity (vgl. 48)
(GENRET)	E8	36	PRE JDE : Adverb/ Preposition Ambiguity

Bei den nicht in Pass 1 behandelten „Wortklassen“ handelt es sich im wesentlichen um Satzzeichen, darüber hinaus um nicht zu definierende Wörter, um mehrdeutige Wörter und um Sonderfälle wie Interjektionen und Partikel.

#### *Beschreibung der einzelnen Programmzweige (PZ) von Pass 1*

*Vorbemerkung:* Was schon für die Beschreibung der Hauptschritte gilt, trifft besonders für die Einzeldarstellung der Programmzweige zu: Die Komplexität des Programms erschwert eine verbale Beschreibung. Es kann hier also nur ein Überblick über die Informationsauswertung und den Ablauf gegeben werden, weitere Einzelheiten sind in den Flussdiagrammen zu diesen Zweigen und zu den in Pass 1 verwendeten (und mit wenigen Ausnahmen nicht näher erläuterten) Unterprogrammen (Subroutines) und Unter-Unterprogrammen (Nested Subroutines) zu verfolgen.

1 PZ FV100 (Finites Verb, POS = 04)

- a) *Prädikat*: Diese Wortklasse wird stets als Prädikat aufgefaßt.
- b) *Infinitivzuordnung*: Falls rechts bereits ein Infinitiv auftrat, wird er, falls das Verb diese Möglichkeit vorsieht, zugeordnet. Befindet sich rechts kein Infinitiv, wird - abweichend von der Grundregel - bis zu acht Wörter (!) nach links ein Infinitiv gesucht und falls vorhanden dem Verb zugeordnet. Etwaige noch nicht zugeordnete SUB rechts des Verbs werden in diesem Fall dem rektionstragenden INF als Objekte angegliedert.
- c) *MNOGO-Zuordnung*: Kann das Verb ein Akkusativ-Objekt regieren, befindet sich links (max. Suche: sechs Wörter) des Verbs ein Adverb vom Typ MNOGO und rechts ein SUB mit GEN-Kasus, so wird beim Verb ein GEN-Objekt zugelassen und darüber hinaus das SUB dem Adverb zugeordnet.
- d) *Objektzuordnung*: Nach der Erkennung einer möglichen Negation (NE unmittelbar vor dem Verb) werden die direkten Objekte in Abhängigkeit von der Valenz des Verbs und von den Kasus der Nominalgruppen rechts des Verbs zu dem Verb gefügt.
- e) *Satztyp*: Anschließend wird versucht, den Satztyp zu klassifizieren. Dies geschieht in Abhängigkeit von etwa vorhandenen Partikeln (BY = Konditionalsatz, LI = Fragesatz).
- f) *Aufzählung*: Sind mehrere Verben im Satz vorhanden, so werden sie auf Übereinstimmung oder Verschiedenheit in Konjugation und Tempus hin überprüft und die entsprechenden Ergebnisse festgehalten.
- g) *Erinnerungsgrenze*: Das Verb bildet eine Hauptgrenze für die Erinnerung an das bisher aufgefundene Inventar des Satzes. Alle Schalter werden gelöscht außer der Erinnerung an das Verb selbst und an noch nicht befriedigte Kasusreaktionen (mögliche Objekte links des Verbs).

Eine Hauptschwierigkeit, die nicht befriedigend gelöst ist, bildet die Objektzuordnung. Es ist möglich, dass ein den Kasusangaben nach mehrdeutiges Nomen (etwa NOM/AKK) als Objekt zugeordnet wird, obgleich es Subjekt ist und das (evtl. sogar eindeutige) Objekt links des Verbs steht. Zwar wird festgehalten, wenn dabei auf Grund der Kasusmehrdeutigkeit ein NOM möglich ist, doch erfolgt später keine Korrektur, falls kein anderes Nomen auftritt, das im Nominativ steht. Treten also mehrere entsprechende Mehrdeutigkeiten auf, wird stets das zuerst gefundene ohne Rücksicht auf Numerusübereinstimmung oder Stellung als Objekt angesehen.

2 PZ SRFAP10 (Kurzform-Adjektiv/Partizip, POS = 08)

- a) *Prädikat*: Falls rechts ein Hilfsverb steht, wird das Ausgangswort diesem zugeordnet. Andernfalls wird es als (Ersatz-)Prädikat gekennzeichnet.
- b) *Objektzuordnung* vgl. die Beschreibung in 1 d
- c) *Satztypbestimmung* vgl. die Beschreibung in 1 e
- d) *Erinnerungsgrenze* vgl. die Beschreibung in 1 g
- Zum Abschluss wird festgehalten, dass ein Kurzform-AD J/PTZ gefunden wurde.

### 3 PZ NETFPR (NET, POS = OG)

- a) *Prädikat*: Es wird notiert, dass NET ein Prädikatsersatz sein kann.
- b) *Einfache Objektzuordnung*: Falls rechts ein noch unbefriedigtes Nomen steht, wird es auf Übereinstimmung mit der Rektion von NET hin überprüft, und falls möglich, als Objekt zugeordnet. Wenn keines vorhanden ist oder keine Zuordnung erfolgen kann, wird festgehalten, dass NET nach links noch ein Objekt binden kann.
- c) *Erinnerungsgrenze*: NET ist Hauptgrenze: die Erinnerung an bereits aufgetretenes Inventar rechts von NET wird vollständig gelöscht.

### 4 PZ NOUN100 (Substantiv, POS = 10, 14, 18)

- a) *Subjekt*: Bei eindeutigem Nominativ wird das Wort als Subjekt gekennzeichnet.
- b) *Objektzuordnung*: Falls rechts ein Verb mit noch freier Rektion vorhanden ist, das SUB kein Subjekt ist und es vom Kasus her dem Verb als Objekt dienen kann, wird es zugeordnet und der Programmzweig verlassen. Steht rechts ein Infinitiv und kann das SUB von einem INF regiert werden, wird es entsprechend zugeordnet.
- c) *Genitivattribut*: Ist rechts ein weiteres SUB vorhanden und steht es im Genitiv, dann wird es als GEN-Attribut dem Ausgangswort zugeordnet. Ist dies nicht möglich, werden Kasusgleichheit oder -verschiedenheit der SUB vermerkt.
- d) *Right Agreement*: Steht rechts ein noch nicht zugeordnetes Adjektiv, wird versucht, es dem SUB als Attribut zuzufügen (Kasus- und Numerusübereinstimmung).
- e) *Erinnerung*: Es wird festgehalten, dass ein SUB im Satz aufgetreten ist. War bereits eines vorhanden, wird auch an dieses vorletzte (in Reserve) erinnert.  
Beim bisherigen Programmstand werden Genitivattribute nicht an links vom Verb stehende Objekte angegliedert (vgl. 4b).

### 5 PZ ABBREV10 (Abkürzung, POS = 1C)

- a) *Genitivattribut*: Es wird notiert, dass dieses Wort kein Genitivattribut sein kann.
- b) *Substantiv*: Anschließend wird der gesamte PZ NOUN100 durchlaufen, das Wort also als Substantiv behandelt.

### 6 PZ ADJPART10 (Adjektiv/Partizip-Langform, POS = 20, 24, 28, 2C; 80, 84, 88, 8C)

- a) *Verneinung*: Das Negationswort NE links vom ADJ wird - falls vorhanden - zugeordnet.
- b) *Attribut bei „Left Agreement“*: Zwei Hauptarten der attributiven Zuordnung sind möglich: Ist bereits vorher eine Zuordnung zwischen einem ADJ und einem SUB erfolgt, wird geprüft,

(1) ob das Ausgangswort diese Zuordnungsgruppe erweitern kann - im Erfüllungsfall wird es zugeordnet und der Programmzweig verlassen (Typ ADJ + ADJ + ... + SUB). Ist dies nicht möglich, wird geprüft,

(2) ob das ADJ als Objekt regieren kann und im Ja-Fall eine entsprechende Zuordnung durchgeführt. Ist dies auch nicht möglich, wird untersucht,

(3) ob rechts eine weitere Nominalgruppe steht, der das ADJ zugeordnet werden kann (Nominale Klammer, Typ ADJ + [ADJ + SUB] + ADJ + SUB). Wenn ja, wird es zugeordnet und der Programmzweig verlassen. Ist rechts noch keine Zuordnung zwischen einem ADJ und einem SUB erfolgt, wird das Ausgangswort bei Deklinationübereinstimmung einem rechts stehenden

(4) SUB zugeordnet, das auch bereits als Objekt erkannt sein kann (Typ ADJ + SUB), oder einer Zuordnungsgruppe

(5) in Reserve, falls eine derartige Gruppe vermerkt ist (Typ ADJ [+ SUB] + ADJ + SUB). Lässt es sich in diesen Fällen nicht zuordnen, wird weiter verfahren wie in (2)f. beschrieben. Wenn keine Zuordnung zu einem SUB nach rechts möglich ist, wird schließlich geprüft, ob rechts ein noch freies ADJ steht, das nach links zugeordnet werden kann (Right Agreement). Ist dies der Fall, kann das Ausgangswort bei Deklinationsübereinstimmung

(6) als Hauptwort mit dem freien ADJ verbunden werden.

c) *Freies ADJ für „Right Agreement“*: In allen Fällen, in denen das ADJ nicht zugeordnet werden konnte, wird festgehalten, dass ein nach rechts nicht zuordnungsfähiges ADJ gefunden wurde, das möglicherweise mit einem SUB links vor ihm verbunden werden kann in einem sog. Right Agreement (Typ SUB + ADJ). Dies wird später z. B. im Programmzweig NOUN100 überprüft.

*7 PZ AD VERB 10 (Adverb, POS = 30)*

a) *Anreihung*: Bei angereihten Adverbien, die auch durch ein Komma getrennt sein können, wird die Anreihung notiert, falls das letzte ADV (rechts) bereits zugeordnet ist.

b) *Verbzueordnung*: Steht unmittelbar rechts vom ADV ein VRB, wird das Ausgangswort diesem VRB zugeordnet (Left Agreement). Dies geschieht auch, wenn unmittelbar rechts ein Kurzform-ADJ/PTZ steht.

*8 PZ ADV3110 (Adverb vom Typ MNOGO, MALO, POS = 34)*

a) *Nominale Zuordnung*: Steht rechts ein SUB oder ein „freies“, d. h. noch nicht zugeordnetes ADJ, wird dieses dem „Adverb“ als „Objekt“ zugeordnet. Zusätzlich wird festgehalten, dass ein ADV vom Typ MNOGO aufgetreten ist.

b) *Verbale Zuordnung*: Ist rechts ein Verb vorhanden, wird das ADV diesem als Attribut (Left Agreement) zugeordnet.

*9 PZ SHFNEUTS (Kurzform-ADJ/PTZ Neutrum Singular, POS = 38)*

a) *Unterscheidung* zwischen ADJ und PTZ wird durchgeführt.

b) *Durchlauf* des Programmzweiges ADVERBIO.

*10 PZ AUX100 (Hilfsverb, Hilfsverb-Infinitiv, POS = 40,44)*

a) *Satztypbestimmung* vgl. 1 e.

b) *Prädikat*: Das Wort wird als Prädikat aufgefasst, falls es Verb ist (POS = 40).

c) *Prädikativum*: Stehen Kurzform-ADJ/PTZ rechts, so werden sie dem Hilfsverb als Prädikativum zugeordnet. Der alte Prädikat-Ersatz-Anzeiger bei dem Prädikativum wird gelöscht. Wenn vorhanden, werden dabei auch angereihte Hilfsverben einbezogen.

d) *Objektzuordnung*: Einfache Zuordnung mit noch freien SUB oder ADJ rechts.

e) *Erinnerungsgrenze*: Wenn das Ausgangswort ein Verb ist (POS = 40), wird die Erinnerung an bereits festgestelltes Inventar rechts gelöscht.

f) *Freie Rektionen*: Sind beim VRB noch freie Rektionen (Objektverlangen) vorhanden, wird dies vermerkt. Zusätzlich wird festgehalten, dass ein Verb aufgetreten ist.

11 PZ INF100 (Infinitiv, POS = 48)

a) Satztypbestimmung vgl. 1 e

Falls Infinitiv noch nicht von rechts erfasst ist (vgl. 1 b)

b) Negation.Zuordnung nach rechts bei noch freien Valenzen des INF.

c) Objektzuordnung nach rechts bei noch freien Valenzen des INF.

d) Vermerk, falls noch freie Valenzen vorhanden sind.

e) Erinnerung, dass ein Infinitiv/Verbalwort aufgetreten ist.

12 PZ GER UND 10 (Gerundium, POS = 4C)

a) Vgl. 11 b-d

b) Vermerk, dass Verbalwort aufgetreten ist.

13 PZ PREP10 (Präposition, POS = 50)

Falls das Wort noch nicht als Governor erkannt ist (dies geschah, wenn rechts ein Relativwort stand)

a) Nominale Zuordnung: Einfache Zuordnung zwischen der PRP und einem - auch bereits erweiterten - SUB rechts. Falls dies möglich ist, wird der Programmzweig verlassen.

b) Erinnerungsgrenze: Die Erinnerung an ein etwa rechts stehendes SUB wird gelöscht.

14 PZ ANOXOTIO (Konjunktionen A, NO, XOT4, POS = 5C)

a) Anreihung: Steht die Konjunktion zwischen zwei zuordnungsfähigen ADJ/PTZ, dann werden die Angaben des rechts stehenden ADJ/PTZ auf die Konjunktion übertragen, d. h. sie wird in die Anreihung einbezogen; anschließend wird ins Hauptprogramm zurückgekehrt. Dies gilt auch für derartig angereihte Adverbien.

b) Erinnerungsgrenze: (Falls a nicht zutrifft) : Alle Schalter werden gelöscht.

c) Angereicherter Nebensatz: Steht vor der Konjunktion ein Komma, wird es als Grenzanzeiger für einen angereichten Nebensatz gekennzeichnet.

15 PZ COMMA10 (Komma, teilweise von I, ILI mitbenutzt, POS = 60)

a) Rückkehr: Falls das Komma bereits von rechts her erkannt wurde, erfolgt gleich die Rückkehr ins Hauptprogramm (HP).

b) Anreihung: Steht das Komma (oder die Konjunktion, für sie gilt in Zukunft das gleiche) zwischen zwei zuordnungsfähigen ADJ/ PTZ oder zwischen zwei ADV, werden die Angaben des rechten Ausgangswortes auf das Komma übertragen und festgehalten, dass seine Funktion erkannt ist. In der gleichen Weise wird bei den rechts und links stehenden Wortklassen SUB, INF und Kurzform-ADJ verfahren. Ist eine derartige Anreihung festgestellt, wird ins HP zurückgekehrt.

c) Gerundiumsatz: Steht links vom Komma ein SUB und ist rechts entweder ein freies SUB, ein freies zuordnungsfähiges ADJ/PTZ oder ein Gerundium vorhanden und findet sich rechts zusätzlich noch ein nicht bestimmtes Komma, dann wird das Ausgangswort als Grenze und Beginn

eines Gerundiumsatzes klassifiziert und das zweite (freie) Komma als Grenze und Ende des Gerundiumsatzes bestimmt.

d) *Vergleich*: Steht rechts das Wort CEM, dann wird festgehalten, dass das Komma vor CEM steht. Seine Funktion ist damit erkannt.

e) *Präpositionalsatz*: Steht rechts eine Präpositionalgruppe, nach der wieder ein (freies, d. h. nicht anderweitig bestimmtes) Komma steht, werden die Satzzeichen entsprechend gekennzeichnet.

f) *Anreihung gleicher Wortklassen*: Stehen rechts und links vom Komma gleiche Wortklassen (außer den in b behandelten), wird festgehalten, dass das Komma anreihende Funktion hat.

g) *Erinnerungsgrenze*: Konnte keine der beschriebenen Kommafunktionen erkannt werden, wird angenommen, dass es eine Satzgrenze darstellt. Daher wird die Erinnerung an noch freie Valenzen, an Infinitive oder Substantive gelöscht. Zum Abschluss wird festgehalten, dass ein noch nicht definiertes Komma (oder Konjunktion) aufgetreten war.

#### 16 PZ IIII10 (Konjunktionen I, III, POS = 64, 68)

a) *Teilsatzgrenze*: Falls links der Konjunktion ein Komma steht, wird das Komma als Satzgrenze gekennzeichnet und die Erinnerung an das bisher aufgetretene Inventar gelöscht.

b) *Fortsetzung*: Falls kein Komma links steht, wird mit dem Programmzweig 15b (Komma) fortgefahren.

#### 17 PZ PER SPRS1 (Personalpronomen im Nominativ, POS = 70)

a) *Subjekt*: Diese Wortklasse wird stets als Subjekt aufgefasst.

#### 18 PZ PRONNSUB (Personalpronomen, das nicht Subjekt sein kann, POS = 74)

a) *Genitivattribut*: Diese Wortklasse kann kein Genitivattribut sein (Kennzeichnung).

b) *Objekt*: Wenn rechts ein VRB mit noch freier Rektion vorhanden ist, wird das Pronomen bei Kasus-Zuordnungsfähigkeit als Objekt bestimmt und ins HP zurückgekehrt.

c) *Erinnerung*: Steht rechts ein SUB, wird es an die zweite Erinnerungsstelle (Reserve) gesetzt und an das Pronomen als erstes Nominalwort erinnert.

d) *Erinnerungsgrenze*: Steht links ein Komma, wird nicht an das Pronomen erinnert, andernfalls wird es als SUB aufgefasst und entsprechend erinnert.

#### 19 PZ EGOEEIX (einschließlich des Unterprogramms (UP) EEEOIS1, Possessivpronomen, POS = 78)

a) *Right Agreement*: Steht rechts eine Präposition oder ein Personalpronomen und links ein SUB, so wird das SUB als Hauptwort eines Right Agreement (Zuordnung rechts) aufgefasst und das Ausgangswort diesem zugeordnet. Ist nur die erste Bedingung erfüllt, wird festgehalten, dass das Ausgangswort in einem Right Agreement stehen kann. Anschließend wird ins HP zurückgekehrt.

b) *Left Agreement*: Falls rechts bereits eine nominale Zuordnung durchgeführt ist, wird versucht, das Possessivpronomen anzugliedern. Ist dies nicht möglich und steht links ein Wort mit noch freier Rektion, die mit dem Kasus des rechts stehenden SUB übereinstimmt, wird eine Objektzuordnung zwischen diesen Wörtern durchgeführt und das Ausgangswort eingegliedert.

c) *Objekt*: Stimmt die Rektion nicht überein und ist sie eine GEN- oder AKK-Rektion, wird das Ausgangswort als Objekt des links stehenden Wortes aufgefasst.

d) *Keine Zuordnung*: Ist keine Zuordnung möglich, wird festgehalten, dass das Possessivpronomen noch in einem Right Agreement stehen kann.

20 PZ KO TOR 10 (einschließlich des UP KO TORS, Relativwort, POS = 7C)

- a) *Subjekt*: Die Wortklasse wird als Subjekt aufgefasst, falls ein Nominativkasus vorhanden ist. Im Ja-Fall Fortsetzung bei d.
- b) *Präpositionalgruppe*: Steht links eine Präposition und ist eine Übereinstimmung im Kasus möglich, wird das Relativwort der Präposition als Satisfier zugeordnet.
- c) *Objekt*: Ist rechts ein VRB vorhanden und wird ein noch freies Kasusverlangen durch das Relativwort erfüllt, wird es als Objekt zugeordnet.
- d) *Relativsatzgrenze*: Suche nach einem Komma links. Das erste gefundene Komma wird als Relativsatzgrenze gekennzeichnet.
- e) *Bezugswort*: Anschließend wird versucht, links vom Komma ein SUB zu finden, das in Genus und Numerus mit dem Relativwort übereinstimmt. Wird es gefunden, dann werden entsprechende Bezugshinweise angebracht.

21 PZ CEM100 (CEM, POS= 90)

Es wird festgehalten, dass ein CEM aufgetreten ist.

22 PZ COMPAR 10 (einschließlich des PZ COMPNSAM, Komparativ - normal, unregelmäßig, mehrdeutig, POS = 94)

- a) *Mehrdeutigkeit*: Komparativ - Positiv Neutrum Singular: falls vorhanden, wird unter e fortgefahren.
- b) *CEM (+ Komma)* : Steht rechts ein CEM, evtl. davor ein Komma, und kein SUB, werden diese Wörter als nach ihrer Funktion (Hinweis auf Komparativ) erkannt gekennzeichnet und es wird ins HP zurückgekehrt.
- c) *CEM (+ SUB)* : Steht rechts CEM und zusätzlich ein SUB, bei dem NOM möglich ist, wird das SUB als Subjekt gekennzeichnet.
- d) *Objekt'-Zuordnung*: Steht rechts ein SUB im GEN, so wird das Ausgangswort mit GEN-Rektion ausgestattet und anschließend eine Objektzuordnung durchgeführt.
- e) *Auflösung der Mehrdeutigkeit*: Steht rechts CEM, wird als Lösung Komparativ angenommen und unter d fortgefahren.
- f) *Adjektiv*: Steht rechts ein SUB im Neutrum (Singular) , dann wird das Ausgangswort als Adjektiv klassifiziert und mit dem Programmzweig ADJPART10 fortgefahren.
- g) *Komparativ*: Steht rechts ein SUB im Genitiv, wird das Ausgangswort als Komparativ betrachtet und nach Ausstattung mit GENRektion eine Objektzuordnung durchgeführt.

23 PZ B Y100 (Konditionalpartikel, POS = 98)

Es wird inventarisiert, dass BY aufgetreten ist.

24 PZ LI100 (Fragepartikel LI, POS = 9C)

Es wird festgehalten, dass LI aufgetreten ist.

25 PZ NUMBER 10 (einschließlich des UP NUMBERS [Zahlen], POS = AO)

- a) *Deklinationsangaben*: Falls die Zahl auf 1, aber nicht auf 11 endet und rechts kein SUB vorhanden ist, werden alle Kasus, Genera und Numeri zugelassen und festgehalten, dass ein SUB gefunden wurde. Rückkehr ins HP.
- b) *Zuordnung*: Steht rechts ein (auch bereits zugeordnetes) SUB, wird eine nominale Zuordnung zum Zahlwort durchgeführt.
- c) *Genitivattribut*: Ist die Zuordnung nicht möglich, wird festgehalten, dass die Zahl kein Genitivattribut sein kann.
- d) *Keine Zuordnung*: War keine Zuordnung möglich, werden in Abhängigkeit von der Art der Zahl einige Deklinationsangaben angefügt und inventarisiert, dass die Zahl substantivisch gebraucht ist.

26 PZ CTO 100 (*Pronomen, POS = BO*)

- a) *Komma*: Falls links kein Komma vorhanden ist, wird ohne Ergebnis ins HP zurückgekehrt.
- b) *Anreihung*: Falls rechts ein weiteres CTO vorhanden ist, werden beide als angereiht gekennzeichnet.
- c) *Erinnerungsgrenze*: Das Komma wird als Grenzanzeiger für eine Teilsatzgrenze klassifiziert und die Erinnerung an das gesamte bisher gefundene Inventar außer an das neue CTO selbst gelöscht.

27 PZ SWCLB 10 (*CTOB Y, KAK, KOGDA, POS = B4, B8, BC*)

Falls links ein Komma steht, wird es als Grenzanzeiger für Teilsätze bestimmt und alle Erinnerungen an aufgefundenes Inventar gelöscht.

### **Anmerkungen:**

1 Vgl. dazu a) PETER TOMA: *The Dictionary Look-up*; b) -, *SERNA-System*; beide: Washington, Georgetown University, 1959.

2 Ca. 100 Seiten; als Anlage zum Arbeitsbericht Nr. 7 der Linguistischen Arbeiten des Germanischen Instituts und des Instituts für Angewandte Mathematik der Universität des Saarlandes, Saarbrücken 1969.

2a Das kyrillische Alphabet der russischen Texte wird vermittels einer Kombination aus Buchstaben des lateinischen Alphabets und Ziffern transliteriert.

3 Die Einführung des Pass 0 mit den beiden Zuordnungen wurde von der Saarbrücker Arbeitsgruppe angeregt.

4 Es lassen sich kaum linguistische Argumente für diese Vorgehensweise finden. Ebenso gut hätte die Analyse hier von rechts nach links erfolgen können. Dabei wäre sogar Pass 0 überflüssig gewesen.

5 Die „alten“ Wortklassenkodierungen, die *allein* im Wörterbuch angegeben sind, wurden bereits im 7090-Verfahren verwendet (vgl. Allgemeine Vorbemerkung 1).