

## 45. Metataxe: ein Dependenzmodell für die computerlinguistische Praxis

1. Grammatikmodellgestaltung als Entwurfsaufgabe
2. Syntax und maschinelle Übersetzung
3. Einsprachige Dependenzsyntax im Modell Metataxe
4. Die eigentliche Metataxe: dependenzsyntaktische Strukturübertragung
5. Modellentwurf und dependenztheoretisches Umfeld
6. Fazit
7. Literatur in Auswahl

### 1. Grammatikmodellgestaltung als Entwurfsaufgabe

Zwei Grammatikmodelle, die in der Lage sind, dieselben empirischen Fakten zu beschreiben, können sehr unterschiedlich gestaltet sein, ohne dass notwendigerweise das eine richtiger wäre als das andere. Auch lässt sich nicht immer a priori entscheiden, ob ein Modell „besser“ ist als ein anderes. Vielmehr ist eine gut begründete Auswahlentscheidung im Falle alternativer Modelle oft sinnvollerweise von der Frage nach dem Zweck bestimmt, dem nach diesem Modell erstellte Grammatiken dienen sollen.

Dieser Beitrag beschreibt eine dependenzuelles Grammatikmodell, das anders ist als andere, weil es einem speziellen Zweck dient: das Modell *Metataxe*, das ich für die maschinelle Übersetzung entwickelt habe. Viele Grammatikmodelle dienen in erster Linie der einsprachigen Deskription und sind statischer Natur. Auch viele Grammatikmodelle, die transformativ oder generativ heißen, sind in dem hier gemeinten Sinne statisch. Ihr generatives Element liegt auf der Theorieebene und dient dazu, durch Regeln aus Axiomen und Grundelementen sprachliche Formen zu erzeugen. Es stellt nicht die Umsetzung tatsächlicher Texte in andere tatsächliche Texte dar. Übersetzungsgrammatiken dagegen beschreiben gerade diesen Vorgang: die Übertragung eines in einer Ausgangssprache formulierten Inhalts in eine Zielsprache. Ich nenne diesen Typ der Grammatik daher dynamisch. Grammatiken, die nicht allein der Übersetzung, sondern spezieller der maschi-

nellen Übersetzung dienen, sind in einem besonderen Sinne auf Dynamik gerichtet, da sie Grundlage eines automatisierten Umsetzungsprozesses sind.

Die maschinelle Übersetzung stellt an eine Grammatik spezifische Anforderungen. Eine Grammatik, die diesem Zweck dient, braucht jedoch nicht selbst als ausführbares Programm direkt in einer Programmiersprache formuliert zu werden. Zu den hier anzuwendenden Grundsätzen gehört als wesentliche Komponente das Prinzip der Modularität. Modularität ist ebenso eine Notwendigkeit professionellen softwaretechnischen und sprachtechnologischen Arbeitens, wie sie auch essentielle Voraussetzung aussagekräftiger und nachvollziehbarer linguistischer Theoriebildung ist. Auf dem Gebiet der maschinellen Übersetzung legt das Prinzip der Modularität es neben vielem anderen nahe, die vier Ebenen (1) des sprachübergreifenden Grammatikmodells, (2) der Grammatik einer bestimmten Sprache, (3) der Formalisierung dieser Grammatik und (4) ihrer Implementierung getrennt zu erarbeiten, zu testen und zu pflegen (vgl. auch Schubert 1988a). In größeren, interdisziplinären Teams ist dies auch deshalb sinnvoll, weil die für die einzelnen Aufgaben erforderlichen Qualifikationen unterschiedlich sind: Modellentwurf und Grammatik sind Sache von Linguisten und Übersetzungswissenschaftlern, die Formalisierung obliegt Computerlinguisten und die Implementierung Softwaretechnologen. Auf die ersten beiden Ebenen gehe ich hier ein; die dritte und vierte liegen im Wesentlichen außerhalb des Rahmens dieses Beitrags (vgl. aber Hellwig 1986; 1989, 358–359; Schubert demn.).

Das hier beschriebene Modell trägt den Namen *Metataxe*. Lucien Tesnière (1959/1982, 283) hat die Bezeichnung *métataxe* für die syntaktische Strukturveränderung eingeführt, die ein Satz bei der Übersetzung erfährt, also etwa für das, was die maschinelle Übersetzung heute *Strukturübertragung* oder *strukturellen Transfer* nennt. Ich verstehe den Begriff durchaus im Tesnièreschen Sinne, verwende ihn jedoch als Etikett für das ganze Grammatikmodell, das sowohl die einsprachige Syntax als auch die Syntax des Über-

setzungsvorgangs (die eigentliche Metataxe) umfasst. Dies ist nicht als Ummünzung des Begriffs *Metataxe* gemeint, sondern als ein Name, der das Typische einer Theorie hervorhebt, die – technischer als Tesnière es meinte – auf die „technique de la traduction“ (Tesnière 1959/1982, 284) und damit auf eine dynamische Anwendung der Grammatik gerichtet ist. (Tesnière 1959/1982, 50–51 verwendet den Gegensatz statisch-dynamisch in einem anderen Sinne; vgl. Bechraoui 1992.)

Das Grammatikmodell Metataxe stammt aus dem Projekt *Distributed Language Translation* (DLT) und ist dort zum ersten Mal dokumentiert (Schubert 1986, 1987). DLT war ein umfangreiches Forschungs- und Entwicklungsprojekt zur maschinellen Übersetzung in weltweiten Kommunikationsnetzen, das in den Jahren 1982 bis 1990 von dem Softwarehaus BSO/Buro voor Systeemontwikkeling BV in Utrecht (Niederlande) betrieben wurde und mehrere funktionierende Prototypen eines maschinellen Übersetzungssystems hervorgebracht hat, die sich des hier beschriebenen Grammatikmodells bedienen (zu Einzelheiten des Projekts vgl. Witkam 1983; Schubert 1988b, demn.; Sadler 1991). Der vorliegende Beitrag beschreibt eine weiterentwickelte Fassung des Modells Metataxe.

## 2. Syntax und maschinelle Übersetzung

Bevor das Grammatikmodell entworfen werden kann, soll ein Blick auf seinen Verwendungszweck klären, welchen Anforderungen es zu genügen hat.

### 2.1. Analyse, Übertragung, Synthese

Die Theorie der maschinellen Übersetzung unterscheidet drei grundlegende Systemarchitekturen: Direkt-, Transfer- und Zwischensprachsysteme. Igor' Mel'čuk nennt diese drei Architekturen morphologisch, syntaktisch und semantisch (Mel'čuk, private Mitteilung 22. 10. 1998), womit die typische, aber nicht die alleinige Operationsebene der jeweiligen Systeme bezeichnet ist: Direktsysteme übersetzen Wörter und feste Wortverbindungen isoliert und erreichen in der Rohübersetzungsqualität meist nicht die Verständlichkeitsschwelle. Transfersysteme übersetzen syntaktisch analysierte (geparste) Sätze. Prak-

tisch alle marktgängigen Systeme, die zumindest minimalen Qualitätsansprüchen gerecht werden, verwenden heute die eine oder andere Art der Transferarchitektur. Zwischensprachsysteme überführen den Inhalt des Ausgangstextes zunächst in eine von Ausgangs- und Zielsprachen unabhängige Zwischenrepräsentation, aus der in jede beliebige Zielsprache weiterübersetzt werden kann. Eine solche sprachunabhängige Zwischenrepräsentation ist ein theoretisches Konstrukt. Meines Erachtens kann kein künstliches Symbolsystem die erforderliche Ausdrucksfähigkeit erreichen, sodass als Zwischenrepräsentation nur eine bereits existierende Sprache in Frage kommt (vgl. Hjelmslev 1963, 101; Schubert 1992a, 83–85). Wählt man jedoch als Zwischenrepräsentation eine *Sprache*, so ist die Systemarchitektur nichts anderes als zwei aufeinander folgende Transferdurchläufe üblicher Art.

Das Grammatikmodell Metataxe ist für das maschinelle Übersetzungssystem DLT entworfen, das eine solche doppelte Transferarchitektur aufwies, wobei als Zwischensprache ein syntaktisch eindeutig gemachtes Esperanto fungierte (Schubert 1988d/1997, 1992a). Das Modell umfasst darum eine einsprachige Syntax, die jeweils der Analyse oder der Synthese dient, sowie eine zweisprachige Übertragungssyntax, die eigentliche Metataxe.

### 2.2. Entscheidungsraum und Entscheidungsmechanismus

Das Grammatikmodell Metataxe ist auf maschinelles Übersetzen gerichtet. Der Entwurf des Modells orientiert sich daher vor allem an der Übersetzungsfunktion.

Eine Grundidee des Modells ist eine möglichst klare Trennung von Form und Inhalt des sprachlichen Zeichens. Das Modell ist rein syntaktischer Art, wobei ich unter *Syntax* nicht nur den Satzbau verstehe, sondern alle Regelmäßigkeiten, die sich ohne Zuhilfenahme des *Inhalts* aus der *Form* des sprachlichen Zeichens erklären lassen (Schubert 1987, 14–16). (Manche Autoren nennen dies *Syntaktik* im Gegensatz zur *Syntax* = Satzbau.) Die Syntax beschreibt Regeln der formalen Kombinierbarkeit sprachlicher Zeichen; zur Syntax in diesem Sinne zählen daher Regelmäßigkeiten der Kombination von Morphemen in Wörtern, Wörtern in Syntagmen, Syntagmen in Gliedsätzen, Gliedsätzen in Sätzen und Sätzen in Texten.

Die Diskussion um die Rolle der Semantik in der Syntax wird immer wieder geführt. Im Falle des Modells Metataxe liegt dem Streben nach semantikfreier Syntax eine bewusste Entwurfsentscheidung zugrunde. Syntaktische Regeln weisen im Allgemeinen klare Voraussetzungen und klare Bedingungen auf. Im Falle prozeduraler Regeln kommen klare Aktionen hinzu, die ausgelöst werden, wenn die Bedingungen erfüllt sind. Syntaktische Regeln basieren auf Ja-nein-Entscheidungen. Falls unterschiedliche Bedingungskonstellationen gleichzeitig erfüllbar sind, erlauben solche Regeln mehrere alternative Lösungen und lassen also Ambiguität zu. Semantische Regeln sind dagegen im Allgemeinen vage und unscharf. Sie basieren auf relativen Entscheidungen („besser als“, „ähnlicher als“ ...). Sie erlauben es daher oft, aus einer Menge angebotener Alternativen eine als die beste auszuwählen, und können in diesem relativen Sinne für Eindeutigkeit sorgen.

Einer computerlinguistischen Formalisierung lassen sich syntaktische Regeln sehr viel einfacher zuführen als semantische, da Softwaresysteme formale (syntaktische) Bedingungen und Aktionen schnell und zuverlässig verarbeiten können, während ihnen – da sie keine denkenden Wesen sind, denen direktes Verstehen des wiedergegebenen Inhalts möglich wäre – Bedingungen und Aktionen auf der Inhaltsseite des sprachlichen Zeichens nur auf dem Umweg über eine simulierende Formalisierung zugänglich sind.

Geht man von der Diagnose aus, dass Syntax verhältnismäßig einfach formalisierbar aber ambig ist, Semantik dagegen unscharf aber zu relativer Eindeutigkeit fähig, so ergibt sich im maschinellen Übersetzungsvorgang eine deutliche Arbeitsteilung zwischen Syntax und Semantik. Die Syntax erkennt und repräsentiert alternative Möglichkeiten, aus denen Mechanismen auswählen, die von semantischem, pragmatischem und außersprachlichem Wissen gesteuert sind. Das Modell Metataxe ist jener Teil dieses Ganzen, der auf der Formseite des sprachlichen Zeichens arbeitet.

Die Syntax erfüllt die Funktion des Bereitstellens von Möglichkeiten im Laufe des maschinellen Übersetzungsvorgangs mehrmals in unterschiedlicher Form. Beim Parsing stellt die einsprachige Syntax die nach den sprachlichen Regelmäßigkeiten zulässigen Strukturen bereit, die die grammatischen Beziehungen zwischen den Elementen eines sprachlichen Ausdrucks beschreiben. Bei dem

zu analysierenden Ausdruck kann es sich um ein komplexes Wort, ein Syntagma, einen Gliedsatz, einen Satz oder einen Text handeln. Die Analysesyntax lässt in vielen Fällen für eine bestimmte Kette sprachlicher Zeichen mehrere unterschiedliche Analysen zu. Sie löst also Ambiguitäten nicht, sondern repräsentiert sie. Wo mehrere Lösungen syntaktisch zulässig sind, liegt die Aufgabe der Auswahl bei Mechanismen, die von Inhaltswissen gesteuert sind. Grundlage dieser Mechanismen ist die Semantik; ihre Formalisierung ist Gegenstand der Wissenstechnik. Wo das Wissen fehlt oder nicht ausreicht, treten in der Praxis der maschinellen Übersetzung willkürliche Auswahlmechanismen in Aktion (wie z. B. „die zuerst gefundene Lösung verwenden“). Die Qualität maschineller Übersetzungen ermisst sich hier wesentlich am Verhältnis zwischen Wissen und Willkür.

In ähnlicher Form funktioniert die zweisprachige Syntax im Strukturübertragungsschritt. Für eine zu übertragende Ausgangssprachliche Struktur kann die Metataxe mehrere, syntaktisch gleichermaßen zulässige Zielsprachliche Äquivalente bereitstellen, wobei die Auswahl wiederum Mechanismen überlassen bleibt, die auf der Inhaltsseite des sprachlichen Zeichens agieren.

Die Frage, ob Auswahlmechanismen erst eingreifen, wenn alle syntaktisch möglichen Alternativlösungen erzeugt und notiert sind, oder ob sie vielmehr schon während des syntaktischen Bearbeitungsschritts aktiv zu werden haben, ist im Wesentlichen computerlinguistischer Natur, da sie erst dann Bedeutung erlangt, wenn es um die Umsetzung der Grammatik in Bearbeitungsprozeduren geht. Die Grammatik selbst ist deklarativ; erst ihre computerlinguistische Formalisierung ist prozeduraler Art.

Eine Syntax nach dem Modell Metataxe hat also die Aufgabe, die Menge der zur Auswahl stehenden alternativen Lösungen zu identifizieren und dadurch den *Entscheidungsraum* festzulegen, in dem dann ein aus inhaltlichen Wissensquellen gespeister *Entscheidungsmechanismus* operieren kann.

### 2.3. Strukturübertragung und lexikale Übertragung

Der Übertragungsschritt ist der Übergang von der Ausgangs- zur Zielsprache. Als solcher ist er Angelpunkt des Übersetzungsvorgangs und Ausgangspunkt des Grammatikmodells, das nach dieser zentralen Phase des Übersetzungsvorgangs benannt ist. Wegen

der in 2.2 beschriebenen Notwendigkeit, formale von inhaltlichen Bearbeitungsschritten zu trennen und ihnen unterschiedliche Funktionen zuzuweisen, konzipiere ich auch die Übertragungsphase des maschinellen Übersetzungsvorgangs in dieser Form. Die Strukturübertragung oder eigentliche Metataxe umfasst die Umsetzung der Struktur ausgangssprachlicher Zeichenketten in zielsprachliche Strukturen. Dieser Bearbeitungsschritt lässt Ambiguität zu und eröffnet damit den Entscheidungsraum für inhaltsgesteuerte Entscheidungsmechanismen.

In sehr enger Verflechtung mit der Strukturübertragung operiert die lexikale Übertragung. Ihr obliegt es, die Wörter zu übertragen. Hier eröffnet das Wörterbuch den Entscheidungsraum, indem es sehr oft für ein ausgangssprachliches Wort eine Reihe zielsprachlicher Übersetzungentsprechungen zur Auswahl anbietet. Die Entscheidungsmechanismen haben auch hier semantisches, pragmatisches und außersprachliches Wissen einzusetzen, wobei bei marktgängigen Systemen bislang die Willkür das Wissen bei weitem überwiegt. Die lexikale Übertragung ist eine der größten ungelösten Aufgaben der maschinellen Übersetzung. Nach heutigem Kenntnisstand darf man annehmen, dass sie nicht vollständig automatisierbar ist, dass jedoch der mögliche Automatisierungsgrad noch lange nicht erreicht ist (Schubert 1997, 24).

Die lexikale Übertragung und ihre Entscheidungsmechanismen stehen außerhalb des Rahmens dieses Beitrags. (Beschreibungen der semantischen Mechanismen, die mit dem Modell Metataxe zusammen entwickelt wurden, geben Sadler 1989 und van Zuijlen 1992; vgl. auch 5.) Die Wortübertragung und die Strukturübertragung stehen jedoch in einer sehr unmittelbaren Wechselbeziehung, da die Wortübertragung die syntaktische Struktur des Syntagmas oder des Satzes über das hinaus verändern kann, was die Strukturübertragung bewirkt. Dies ist dann der Fall, wenn die syntaktische Struktur übersetzungsäquivalenter lexikaler Einheiten nicht in beiden Sprachen gleich ist (vgl. Schubert 1996, 101–104). Hier greifen die Strukturübertragung und jener grammatikalisierte Wortartwechsel ineinander, den Tesnière (1959/1982, 361) *translation* nennt (vgl. Werner 1993, 111).

#### 2.4. Isomorphie oder Implizität?

Für ein auf maschinelles Übersetzen gerichtetes Syntaxmodell ist die Strukturübertragung der naheliegendste Ausgangspunkt. Hat der

linguistische Systementwurf festgelegt, wie der Übertragungsschritt erfolgen soll, so können sich die Eigenschaften und Funktionen der vor und nach diesem zentralen Bearbeitungsschritt liegenden Systemkomponenten hieran orientieren.

Zwei Typen von Lösungen sind denkbar, jede mit einer Vielzahl möglicher Ausgestaltungen. Der eine Lösungstyp ist der der isomorphen Syntaxen, der andere beruht auf dem Implizitätsprinzip. In 2.1 war von der Systemarchitektur der Zwischensprachsysteme die Rede und davon, dass es die hierfür erforderliche Zwischensprache nicht gibt, es sei denn in der Form einer existierenden Sprache.

Während eine echte Zwischensprache über eigene, von Ausgangs- und Zielsprachen unabhängige strukturelle und lexikale Elemente zu verfügen hätte, ist auch versucht worden, eine hybride Lösung zu verwirklichen, die auf Isomorphie der Syntaxen beruht. Hierbei entstammen die lexikalischen Elemente jeweils den Ausgangs- und Zielsprachen, die Analyse des Ausgangssatzes wird jedoch so weit fortgeführt, bis eine syntaktische Struktur entsteht, die für übersetzungsäquivalente Sätze aller am System beteiligten Sprachen übereinstimmt. Ein solcher Ansatz setzt voraus, dass die Syntaxen der Ausgangs- und Zielsprachen auf einen gemeinsamen Nenner gebracht werden können, der dann auf der Ebene der Struktur als Zwischenglied fungiert. Das Projekt Rosetta ging von einem solchen Ansatz isomorpher Grammatiken aus (Landsbergen 1987, 1994; Hutchins 1988, 38–39). Ein wesentliches Problem dieses Ansatzes lag darin, dass die isomorph gemachten Grammatiken aller beteiligten Sprachen jedes Mal überarbeitet werden mussten, wenn eine weitere Sprache hinzugefügt werden sollte. Dieser theoretisch durchaus viel versprechende Ansatz, bei dem die Satzstruktur von ihrer lexikalischen Füllung getrennt behandelt wurde, hat also die für eine echte Zwischenrepräsentation erforderliche Unabhängigkeit von Ausgangs- und Zielsprachen nicht erreicht.

Das Grammatikmodell Metataxe geht einen anderen Weg. Sein Ausgangspunkt ist ein Grundsatz, den ich Implizitätsprinzip nenne (Schubert 1988c). In der kontrastiven Grammatik, der grammatisch orientierten Übersetzungswissenschaft und speziell in Arbeiten zur Strukturübertragung ist oft davon die Rede, dass in bestimmten Fällen ein Satz bei der Übertragung aus einer Sprache in eine andere einer Strukturumformung unterzogen

werden müsse (vgl. etwa die Beispiele bei Koch 1996, 211). Diese Sichtweise setzt stillschweigend voraus, dass die Syntax von Ausgangs- und Zielsprache weitgehend gleich ist und Umformungen nur in Sonderfällen notwendig werden. Das Modell Metataxe fußt demgegenüber auf der Idee, dass die Struktur jedes einzelnen Satzes (wobei der Satz hier immer stellvertretend für andere mögliche Übersetzungseinheiten steht) bei der Strukturübertragung vollständig zerlegt und in der Zielsprache neu wieder aufgebaut wird.

Die eigentliche Metataxe ist also ein Regelsystem zur *vollständigen Abbildung* aller zulässigen ausgangssprachlichen Strukturen auf zielsprachliche. Sie heißt implizit, weil ihre Regeln sprachliche Formen auf sprachliche Formen abbilden, ohne dabei deren Inhalt in Form einer Metarepräsentation wiederzugeben. Das Aufstellen einer Metataxeregeln setzt voraus, dass Ausgangs- und Zielstruktur bedeutungsgleiche Übersetzungsentsprechungen sind. Die Übersetzungsäquivalenz wird hier durch den Autor der Regel festgestellt, während das maschinelle Übersetzungssystem ausschließlich mit Regeln auf der Formseite des sprachlichen Zeichens arbeitet. Andere Ansätze bemühen sich darum, den Inhalt der Übersetzungseinheiten in einer semantischen Repräsentation, einem logischen Formalismus oder in anderer Weise niederzuschreiben, und machen ihn dadurch explizit.

Der implizite Ansatz ist für die Arbeit der Grammatikentwerfer auch deswegen vorteilhaft, weil er es erlaubt, die einsprachigen Syntaxen der Ausgangs- und Zielsprachen in vollständiger Unabhängigkeit voneinander zu gestalten. Selbst wenn Wortarten und Dependenztypen (vgl. 3.2) oft in den Syntaxen verschiedener Sprachen gleich lautende, traditionelle Namen (wie *Adjektiv* oder *direktes Objekt*) erhalten, so liegt hierin noch keinerlei Festlegung darauf, dass einer so benannten ausgangssprachlichen Struktur in einer bestimmten Zielsprache eine ebenso benannte Struktur zu entsprechen habe. Wollte man dies voraussetzen, so bekäme man auf einer anderen Ebene ganz ähnliche Synchronisierungsprobleme wie bei einem Isomorphieansatz.

### 3. Einsprachige Dependenzsyntax im Modell Metataxe

Für den Menschen ist die Symbolkette, die wir Text nennen, Träger eines Inhalts. Wie kann ein automatisierter Übersetzungsvor-

gang so gestaltet werden, dass der implizite Inhalt erhalten bleibt? Kazimierz Ajdukiewicz weist in seinem klassischen Aufsatz darauf hin, was sinnvolle von sinnlosen Symbolketten unterscheidet. Er stellt fest, dass eine Wortkette nur dann einen Sinn hat, wenn sie den Regeln der Syntax entspricht und somit, wie er es nennt, „syntaktisch konnex“ ist (Ajdukiewicz 1935, 1; vgl. Schubert 1987, 51). Auch für Tesnière (1959/1982, 11) steht der Begriff der „connexion“ zentral. Er macht jedoch auch deutlich, dass es keine Eins-zu-eins-Entsprechung zwischen einem bestimmten Konnexionsmuster (d. h. einer syntaktischen Struktur) und einem bestimmten Inhalt gibt, sondern dass vielmehr derselbe Inhalt in verschiedenen Sprachen durch unterschiedliche Strukturen repräsentiert werden kann. Übersetzen macht also eine Strukturumformung notwendig. Diese nennt Tesnière (1959/1982, 283) *métataxe*.

Indem Tesnière postuliert, dass übersetzungsäquivalente Sätze zweier Sprachen einander semantisch entsprechen, folgt er dem, was ich Implizitätsprinzip nenne. Tesnière bleibt also im Grunde nicht, wie Peter Koch und Thomas Krefeld formulieren, „die – eigentlich semantischen – Tertia comparationis schuldig“ (Koch/Krefeld 1991, 8), sondern er hält sie implizit. Engel (1980, 11) und Koch (1996, 212) weisen auf Unterschiede in der semantischen Analyse übersetzungsäquivalenter Sätze hin, wobei allerdings nicht deutlich wird, ob diese Unterschiede in der Sprache oder im semantischen Analysemodell begründet sind.

Mit dem nach diesem Begriff Tesnières benannten Grammatikmodell schließe ich mich seiner Darstellung in großen Zügen an, arbeite das Modell jedoch mit einer Stringenz aus, wie sie für die computerlinguistische Anwendung erforderlich ist.

#### 3.1. Einheiten und Relationen des dependenzsyntaktischen Systems

In der für Dependenzgrammatiken typischen Weise klassifiziert und benennt das Modell Metataxe als primäre Größe nicht Gruppen zusammengehöriger Wörter, sondern die syntaktischen Beziehungen, die ihre Zusammengehörigkeit ausmachen (vgl. Erosms 1985, 307). Es richtet sich so primär auf die syntaktische Funktion und nutzt die syntaktische Form nur sekundär als Identifizierungsmerkmal (vgl. aber Hellwig 1989: 359; siehe auch 3.4). So schafft das Modell eine Struktur, die übersetzungsrelevante Eigenschaften reprä-

sentiert und die überdies die für die außerhalb der Syntax liegenden Entscheidungsmechanismen erforderliche Verbindung zur Semantik in recht direkter Weise ermöglicht (vgl. 5). Die so definierte syntaktische Struktur ist zwar mit einer auf sie aufsetzenden semantischen Repräsentation nicht deckungsgleich, ist aber unaufwendig in sie überführbar. Auf diesen, auch in Ajdukiewicz's Gedankengang angelegten Übergang zur Semantik komme ich in 5. zurück.

Ein System besteht aus Elementen und den Relationen zwischen ihnen. Für das Modell Metataxe ist das Wort eine naheliegende Einheit. Wenn in Sprachen, die wie etwa Japanisch und Chinesisch eine ununterbrochene Aneinanderreihung der Schriftzeichen erlauben, ebenfalls mit Wörtern gearbeitet werden soll, muss zunächst ein Wortgrenzenerkennungsalgorithmus vorgeschaltet werden. In Sprachen mit komplexen Wörtern, insbesondere in agglutinierenden Sprachen, kann es sinnvoll sein, kleinere Einheiten als das Wort zu wählen, zum Beispiel Morpheme oder feste Morphemgruppen. Von den einsprachigen Dependenzsyntaxen, die nach den Vorgaben des Modells Metataxe erarbeitet worden sind, haben die Autoren der deutschen (Lobin 1989), der dänischen (Ingrid Schubert 1989), der polnischen (Świdziński 1989), der bengalischen (Dasgupta 1989) und der finnischen (Tarvainen 1989) Syntax sowie der des Esperanto (Schubert 1989b) als Einheit das Wort, die der ungarischen (Prószéky/Koutny/Wacha 1989) dagegen das Morphem als Einheit gewählt. Die japanische Arbeit (Sato 1989) zeigt ein Beispiel einer Syntax, die erst nach Wortgrenzerkennung angewendet werden kann. (Weitere Dependenzsyntaxen aus dieser Reihe sind unveröffentlicht geblieben: Englisch: Dan Maxwell; Englisch: Bieke van der Korst; Isländisch: Stefán Briem; Französisch: Luc Isaac; Französisch: Dorine Tamis; Italienisch: Germano Proverbio/Laura Sciollo; Tschechisch: František Čermák; Chinesisch: Li Wei; Arabisch: Mauro Tosco.)

Als grundlegende Relation zwischen den Wörtern setze ich die *Dependenz* an, die ich mit Engel (1977/1994, 28) als „gerichtete Konkomitanz“ beziehungsweise als „directed co-occurrence“ (Schubert 1987, 29) definiere. Richard Hudson (1984, 76) spricht von „companion-ship“ und beruft sich in der Frage der Gerichtetheit dieser Beziehung auf Tesnière. In einer solchen Dependenzbeziehung stehen zwei Wörter, wenn das eine durch seine inhärenten Kombinationseigenschaften

dem anderen das Vorkommen in dem betreffenden Satz ermöglicht (Baumgärtner 1970, 54). Wenn Jürgen Erich Schmidt dagegen meint, Louis Hjelmslev habe „den Dependenzbegriff als undefinierbar erklärt“ (Schmidt 1991, 212 und Anm. 11), so erscheint mir dies missverständlich. Der Terminus *afhængighed* 'Abhängigkeit', den Hjelmslev (1943/1966, 27) als Axiom annimmt, meint sehr viel mehr als syntaktische Dependenz (vgl. Rasmussen 1992, 428–436). Die Dependenzrelation des Modells Metataxe ist jedoch eine *syntaktische* Beziehung. Es wird bisweilen erörtert, inwieweit sie auf logisch-semantische Kombinationsregelmäßigkeiten zurückzuführen ist (vgl. etwa Tarvainen 1981, 16), jedoch ist eine solche semantische Verankerung keine Besonderheit der Dependenz, sondern ein Niederschlag jener sehr viel weiter reichenden Beobachtung, wonach Syntax zu einem festen Regelsystem erstarrte Semantik ist.

In dem durch eine Dependenzrelation verbundenen Wortpaar heißt das Wort, das die Vorkommensstelle eröffnet, *Regens*, während das Wort, das diese Stelle einnimmt, das *Dependens* ist (Abb. 45.1). Ein Wort kann mehrere Dependents, aber nur ein unmittelbares Regens haben. Syntaktisch verbundene Wörter bilden ein *Syntagma*. Das Wort, das innerhalb eines Syntagmas alle anderen Wörter direkt oder indirekt regiert, nenne ich *inneres Regens* des Syntagmas. Wenn das Syntagma als Ganzes von einem weiteren Wort abhängt, so ist dieses in Bezug auf das betreffende Syntagma ein *äußeres Regens* (Abb. 45.2). (Die Bedingung, wonach ein Wort nicht mehr als ein unmittelbares Regens haben kann, ist eine willkürliche Festlegung. Sie ist Bestandteil vieler dependenzgrammatischer Modelle, vgl. z. B. Iomdin 1990, 23. Mel'čuk 1988, 23–24 gibt eine präzise formale Begründung für diese Bedingung. Die Bedingung hat zur Folge, dass die entstehenden Strukturen im Sinne der mathematischen Graphentheorie Baumstrukturen sind, was bei der Formalisierung der Metataxe die Na-

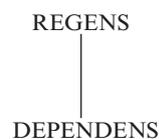


Abb. 45.1: Regens und Dependens

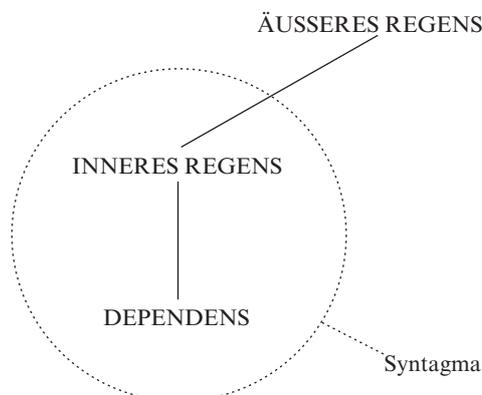


Abb. 45.2: Inneres und äußeres Regens

vigation in der komponentenweise abzuarbeitenden Struktur vereinfacht, vgl. van Zuijlen 1989.)

Tesnière äußert sich nicht allzu ausführlich zu der Frage, wie eine Dependenzbeziehung zu erkennen ist. Wo vorhanden scheint er vor allem die syntaktische Rektion, also die Bestimmung der morphologischen Form des Dependens durch das Regens, als Kriterium zu verwenden, wobei er in Fällen, in denen morphologische Formen fehlen, die durch Rektion bestimmt sein könnten, oft in Analogie zu Fällen (und Sprachen) vorgeht, in denen es einen solchen sichtbaren Hinweis gibt (vgl. Schubert 1987, 41–43). Im Vergleich typologisch unterschiedlicher Sprachen ist dieses Vorgehen problematisch. Die Kriterien Rektion und Analogie (zu indogermanischen Sprachen) führen bisweilen zu Widersprüchen, etwa dort, wo das aus Analogiegründen als Regens eingestufte Wort die morphologische Markierung trägt (vgl. Nichols 1986). In der Frage der Richtung der Dependenzbeziehung kann zwar die syntaktische Rektion als Kriterium dienen, weil sie eine gerichtete Relation ist, in Fällen der Kongruenz jedoch, die ja Rektion in zwei Richtungen ist, reicht dieses Kriterium nicht aus. Engel (1977/1994, 28) nimmt Fälle syntaktischer Kongruenz zum Anlass, darauf hinzuweisen, dass die Frage der Richtung der Dependenzbeziehung der willkürlichen Entscheidung des Grammatikers unterliege. Ich glaube zwar, dass zum Gestalten eines Grammatikmodells oft willkürliche Entscheidungen notwendig sind, denke jedoch, dass sich im vorliegenden Fall ein recht gutes Kriterium in der Beobachtung findet, dass die Kombinationseigenschaften eines Syntagmas im Wesentlichen denen sei-

nes inneren Regens entsprechen (Hudson 1984, 77; Schubert 1987, 36). Durch dieses Kriterium lässt sich das innere Regens in jedem Falle ermitteln, auch dort, wo syntaktische Rektion fehlt oder wo Kongruenz ein Interdependenzverhältnis nahe zu legen scheint.

### 3.2. Wortarten und Dependenztypen

Im Prinzip hat eine Syntax der hier benötigten Art die Aufgabe, zu jedem Wort einer Sprache anzugeben, mit welchen anderen Wörtern es eine syntaktische Relation eingehen kann. Um einer Syntax jedoch – neben praktischen Eigenschaften wie Überschaubarkeit und Formalisierbarkeit – eine über das Aufzählbare hinausgehende vorhersagende Gültigkeit zu verleihen, fasst man sowohl die Wörter als auch die Relationen zu Klassen zusammen und formuliert die Aussagen der Syntax nicht als Angaben zu einzelnen Wörtern, sondern als Regeln über eine überschaubare Zahl von Wortklassen und Relationsklassen.

Bei den Wörtern bietet sich in europäischen Sprachen eine Einteilung in etwa zehn Klassen an. Ich habe gezeigt (Schubert 1987, 46–50), wie Wortklassen distributionell durch Beobachtung paradigmatischer Verhältnisse definiert werden können. Das Ergebnis dieses Verfahrens ist im Allgemeinen eine Einteilung, die der traditionellen Wortarteneinteilung weitgehend entspricht. Sie umfasst die Inhaltswortarten Verb, Adjektiv, Adverb und Substantiv, kommt bei den Funktionswortarten jedoch hier und da zu weniger traditionellen Lösungen, vor allem dort, wo die traditionellen Wortarten, etwa im Bereich der Konjunktionen und Subjunktionen oder der Präpositionen und Postpositionen, distributionellen Kriterien nicht immer genügen. Bei der Einteilung des Wortschatzes in Wortarten hat der Grammatiker eine gewisse Freiheit zu zweckmäßigen und letztlich willkürlichen Entscheidungen. Ist man sehr streng darauf bedacht, dass alle Mitglieder einer Wortart genau dieselben Kombinationseigenschaften aufzuweisen haben, dann erhält man viele Wortarten mit zum Teil nur wenigen Mitgliedern. Lässt man dagegen gewisse Eigenheiten bei Unterwortarten zu, dann kann man mit einer recht traditionellen Wortarteinteilung arbeiten (vgl. 3.5).

Wählt man als Grundeinheit nicht das Wort sondern beispielsweise das Morphem (vgl. 3.1), so ergibt sich eine Einteilung, die

dem über Wörter Gesagten recht nahe kommt: Die drei beziehungsweise vier Inhaltsmorphemklassen (Verb, Adjektiv/Adverb, Substantiv) sind recht deutlich von anderen unterschieden; bei den Funktionsmorphemklassen ist eine Neufestlegung zweckmäßig. Im folgenden nenne ich die Grundeinheit der Einfachheit halber stets *Wort*.

Führt man eine distributionelle Wortklassifizierung in der erwähnten Art und Weise als Paradigmenbildung (Schubert 1987, 46–50) durch, so zeigt sich in vielen Fällen, dass die mit einer bestimmten Wortart in syntaktischer Relation stehenden Wörter mehrere diskrete Klassen paradigmatisch austauschbarer Dependientien und Regentien bilden. Ich wähle als Betrachtungsrichtung die vom Regens zum Dependens und nenne die so entstandenen Klassen *Dependenztypen*. Die Dependenztypen erhalten Namen, die sich soweit möglich an traditionellen Benennungen orientieren: Subjekt, direktes Objekt, indirektes Objekt, freie Angabe usw. (ausführlich Schubert 1987, 51–57). In der grafischen Darstellung erscheinen die Wörter auf den Knoten der Baumstruktur, die Dependenztypen dagegen als Etiketten an den Ästen.

Eine solche distributionelle Neuklassifizierung der syntaktischen Beziehungen ist schon früher mehrfach vorgeschlagen worden. Einer der ersten, die hier rigoros neu definieren, ist Mel'čuk (1964, 18–27) mit seinen *otnošenijsa neposredstvennoj dominacii* 'unmittelbaren Dominanzbeziehungen'.

### 3.3. Rekursivität

Die menschliche Sprache bildet eine unendliche Fülle an Formen und Strukturen. Ein wesentliches Ziel bei der Gestaltung eines Grammatikmodells ist es, ein Regelwerk zu schaffen, das über das Aufzählbare hinausgeht und in der Lage ist, das Regelmäßige in der Unendlichkeit der sprachlichen Produktion zu erfassen. Es ist bekannt, dass die Extrapolation von endlich vielen Beobachtungen zur unendlichen Vielfalt des Möglichen am besten gelingt, wenn man sich bemüht, ein möglichst elegantes Regelsystem zu schaffen. Ein Regelwerk kann dann als elegant gelten, wenn es möglichst wenige, möglichst einfache Regeln enthält, die jeweils eine möglichst kleine, aber häufig vorkommende Teilstruktur der Menge der zu beschreibenden sprachlichen Strukturen wiedergibt. Natürlich muss ein solches Regelwerk die Komplexität der menschlichen Sprache wiedergeben können. Die Komplexität sollte jedoch nach

Möglichkeit nicht in der einzelnen Regel liegen, sondern aus der (oft rekursiven) Kombination einfacher Elementarregeln entstehen.

Das Modell Metataxe bildet syntaktische Gegebenheiten in Form von Baumstrukturen ab. Die einem Satz (oder einem Textstück anderer Länge) zugeordnete Baumstruktur ist definiert durch die Kombinationseigenschaften der Wörter. Die in 3.1 gegebene Verallgemeinerung der Darstellung dieser Kombinationseigenschaften beruht auf Klassenbildung bei den Wörtern und den syntaktischen Relationen in Form von Wortarten und Dependenztypen. Dies erlaubt einen sehr einfachen Aufbau einer Basissyntax, die dann, wo nötig, genauer justiert werden muss.

Die Grundstrukturen der Syntax lassen sich in zwei Typen einfacher Regeln darstellen. Regeln des ersten Typs geben an, welche Dependenztypen von einem Regens einer bestimmten Wortart abhängen können (Abb. 45.3). Regeln des zweiten Typs beschreiben die syntaktische Form der Dependientien eines bestimmten Dependenztyps (Abb. 45.4). Zum ersten Regeltyp gehört beispielsweise eine Regel, die angibt, dass im Deutschen von der Wortart Verb der Dependenztyp direktes Objekt abhängen kann. Zum zweiten Regeltyp gehört eine Regel, die beschreibt, dass ein direktes Objekt einer der Wortarten Substantiv, Pronomen, Verb (als inneres Regens eines Nebensatzes) usw. angehören kann. Indem man diese beiden Regeln getrennt formuliert, statt eine zusammenfassende Regel über das direkte Objekt von Verben festzulegen, vermeidet man es, redundante Regeln aufzustellen. Auch andere Regentien wie zum Beispiel Partizipien können ein direktes Objekt regieren, das genau dieselbe syntaktische Form

REGENS

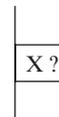


Abb. 45.3: Grundregeltyp 1



X ?

Abb. 45.4: Grundregeltyp 2

hat wie das direkte Objekt eines Verbs. Setzt man für das Verbobjekt eine einzige Regel an, müsste man einen Teil der Information in der Regel über das Objekt des Partizips wiederholen.

Stellt man für eine bestimmte Wortart als Regens alle Regeln des ersten Typs in einer einzigen Baumstruktur zusammen, so ergibt sich eine Darstellung, die die maximale Regenskapazität der betreffenden Wortart zeigt. Eine solche Struktur nenne ich *Elementarbaum*. Ein Elementarbaum besteht aus einem Knoten mit vielen abhängenden Ästen. Jeder Ast trägt sein Dependenztypenetikett. Die Regeln des zweiten Typs geben an, welche syntaktische Form die Dependenztypen jedes einzelnen Dependenztyps annehmen können. Setzt man an der Stelle der Dependenzknoten des Elementarbaums jeweils die Elementarbäume für die dort möglichen Wortarten ein, so erhält man eine Reihe komplexer Bäume. Führt man diesen Vorgang rekursiv aus, so entsteht die Maximalsyntax der jeweiligen Sprache, also eine Beschreibung aller zulässigen Sätze dieser Sprache. Eine solche Maximalsyntax lässt sich grafisch nicht vollständig darstellen, da sie unendlich ist. Als Gedankenexperiment zeigt die Möglichkeit der Erzeugung von Maximalsyntaxen durch rekursives Verknüpfen von Elementarbäumen jedoch, wie ein unendliches System aus wenigen einfachen Regeln erzeugt werden kann.

Zwischen der Dependenzgrammatik und der von Aravind Joshi vorgeschlagenen Tree Adjoining Grammar (vgl. Joshi/Schabes 1991; Stone/Doran 1997 und viele weitere Veröffentlichungen) besteht bekanntlich eine deutliche Verwandtschaft. Im Falle des Modells Metataxe wird dies an dem hier dargestellten Gedankenexperiment besonders deutlich.

Die in diesen beiden Regeltypen beschreibbare Syntax bedarf jedoch der Verfeinerung und Justierung. Im bisher beschriebenen Zustand ist sie das, was die generative Grammatik eine übergenerierende Syntax nennt.

Die beiden genannten Regeltypen beschreiben den kleinsten denkbaren Ausschnitt aus einer komplexeren Struktur, in dem noch ein Element des Systems und eine Relation vorhanden sind. Je kleiner der gewählte Ausschnitt, desto häufiger ist die Regel anwendbar und desto eleganter wird das Regelwerk. Es gibt jedoch Regelmäßigkeiten, die eine größere Zahl an Wörtern und Dependenztypen einbeziehen. Das Verhältnis zwischen einem deutschen Verb und seinem di-

rekten Objekt kann man zwar wie oben gezeigt in zwei Regeln erfassen, von denen die eine angibt, dass ein Verb ein direktes Objekt haben kann (erster Regeltyp), und die andere, dass das direkte Objekt z. B. ein Substantiv sein kann. Letztere Regel sollte genauer gefasst werden, denn über die Wortart hinaus wird noch ein Merkmal durch die Funktion des Wortes als direktes Objekt bestimmt: der Kasus. Welcher der Kasus Akkusativ, Dativ und Genitiv es ist, lässt sich in dem kleinen Ausschnitt aus Dependenztyp und Dependenz nicht ermitteln. Die Kasuswahl ist vom Regens gesteuert. (Diese Beschreibung setzt voraus, dass man Genitive und Dative ebenso als direkte Objekte einordnet wie Akkusative, was mir aus der Sicht des Modells Metataxe sinnvoll erscheint. Lobin (1989), der ebenfalls nach den Vorgaben dieses Modells gearbeitet hat, hat sich anders entschieden und in Engelscher Tradition die Objektdependenztypen nach ihrer Kasusform unterschiedlich benannt. Beide Lösungen sind möglich; meine verursacht etwas größeren Aufwand in der Analysephase, Lobins in der Übertragungsphase.) Wenn die Kasuswahl beim direkten Objekt von einer speziellen Eigenschaft bestimmter regierender Verben abhängt, dann braucht man eine Informationsquelle, die verzeichnet, welches Verb welchen Kasus verlangt. Ich komme bei den Themen Valenz (3.5) und Wörterbuch (3.6) hierauf zurück. Festzuhalten bleibt, dass hier eine Regel vorliegt, die komplexer ist als die beiden Minimalregeltypen.

Eine andere Form der Verfeinerung der durch die Minimalregeltypen konstituierten Basissyntax wird dann notwendig, wenn nicht, wie bisher angenommen, Regentien, die derselben Wortart angehören, auch dieselben Dependenztypen regieren können. Dies kann zum einen daran liegen, dass man die Wortarten so weit gefasst hat, dass es dependenziell unterscheidbare Unterwortarten gibt (vgl. 3.5). In diesem Fall haben also verschiedene Wörter derselben Wortart unterschiedliche Regenseigenschaften. Es gibt jedoch auch den Fall, dass sich die Regenseigenschaften ein und desselben Wortes durch das Hinzutreten syntaktischer Merkmale ändern. Beispielsweise sind im Deutschen die Regenseigenschaften eines Verbs bei Anwesenheit des Merkmals *infinit* anders, als wenn das Merkmal *finit* steht, und innerhalb der Menge der Verben mit dem Merkmal *finit* sind die Regenseigenschaften unterschiedlich je nachdem, ob das Diathesemerkmals Aktiv

oder Passiv steht. (Anders als Badia 1993, 117 glaube ich nicht, dass man zur Erklärung der diatheseabhängigen Regenseigenschaften eine Tiefenstruktur mit semantischen Rollen heranziehen muss.)

#### 3.4. Direkt und indirekt übersetzbare Merkmale

Aus der Festlegung, dass das Grammatikmodell Metataxe alles umfassen soll, was auf der Formseite des sprachlichen Zeichens liegt (vgl. 2.2), ergibt sich, dass es auch die syntaktischen Relationen zwischen den Elementen des Wortes einschließt, also das, was traditionellerweise Morphologie und Wortbildung heißt. Von der Wortbildung kann das Modell jedoch nur den auf der Formseite des sprachlichen Zeichens fassbaren Teil einschließen. Eine Dependenzsemantik des Wortes habe ich anderweitig dargestellt (Schubert 1993).

Das dependenzielle Kombinationsverhalten von Wörtern wird von einigen ihrer *Merkmale* beeinflusst. Zu den Merkmalen zähle ich hier alle syntaktischen und semantischen Eigenschaften, die an den Wörtern einer bestimmten Sprache in grammatikalisierter Form ausgedrückt werden. Als grammatikalisiert können solche Eigenschaften gelten, bei denen eine bestimmte Form- oder Inhaltskategorie eine feste Zahl möglicher Werte annehmen kann, die bei einer großen Zahl von Wörtern in gleichartiger Weise ausgedrückt sind. Hierzu zählen die Kategorien Numerus, Finitheit, Tempus, Modus, Aspekt, Diathese und andere. Ein Merkmal kann durch ein Morphem, die Veränderung eines Stammmorphems, ein Funktionswort oder durch die Wortfolge angezeigt werden.

Welchen konkreten Ausdruck ein bestimmtes Merkmal hat und welche Ambiguitäten bei der Zuordnung von Ausdrucksformen zu Merkmalen bestehen, ist Sache des Parsings, dessen Mechanismen durch Formalisierung einer einsprachigen Syntax entstehen. Es gibt jedoch eine Unterscheidung, die in der Übertragungsphase des maschinellen Übersetzungsvorgangs unmittelbar funktional ist. Dies ist die Unterscheidung zwischen *direkt* und *indirekt übersetzbaren Merkmalen*. Das Parsing hat die Aufgabe, einen linearen Satz in eine Baumstruktur umzusetzen, die die Dependenzrelationen zwischen den Wörtern des Satzes in der hier beschriebenen Weise angibt. Die zum Aufbau dieser Struktur erforderliche Information kommt aus zwei Quellen: (1) Regeln, die Schlüsse aus den

*Merkmale* auf die Dependenzrelationen des betreffenden Wortes zu anderen Wörtern erlauben, und (2) Information, die dem Wort in einem *Wörterbuch* oder einem System von Redundanzregeln zugeordnet ist. Welche Informationsquelle welche Information liefert, ist einzelsprachspezifisch unterschiedlich. Ebenso ist je nach Sprache verschieden, in welcher Art die jeweiligen Merkmale ausgedrückt sind: was in der einen Sprache an der Wortfolge abzulesen ist, wird in der anderen durch Flexion wiedergegeben, und was in der einen Sprache aus der Wortform ableitbar ist, muss in der anderen Sprache für jedes in Frage kommende Wort im Wörterbuch verzeichnet werden. In Sprachen, deren Wörter Merkmale tragen, die Aussagen über die Dependenzrelationen im Satz erlauben, lassen sich meist Merkmale unterscheiden, deren einzige Funktion im Übersetzungsvorgang das Bereitstellen der Dependenzinformation für das Parsing ist. Diese Morpheme nenne ich *indirekt übersetzbar*. Daneben gibt es Merkmale, die Bedeutung tragen. Diese nenne ich *direkt übersetzbar*. (Mel'čuk 1988, 20 spricht in ähnlicher Weise von direktem und indirektem Bezug zur Semantik.)

Ein Beispiel für indirekt übersetzbare Merkmale sind die Kasusmerkmale in den Sätzen *Er braucht meine Hilfe* und *Er bedarf meiner Hilfe*. Ich ordne das Wort *Hilfe* in beiden Sätzen als direktes Objekt des jeweiligen Verbs ein. Die Kasusmerkmale haben nur die Funktion, zusammen mit regenseitiger Information wie der Valenz des Verbs dem Syntagma *meine(r) Hilfe* den Dependenztyp direktes Objekt zuzuweisen. Sobald dies erfolgt ist, ist die Information, dass das Substantiv in dem einen Fall im Akkusativ und in dem anderen Fall im Genitiv stand, nicht mehr erforderlich. Im weiteren Verlauf des Übersetzungsvorgangs verwendet die lexikale Übertragung die direkt übersetzbaren Merkmale (z. B. das Numerusmerkmal), während die Strukturübertragung den Dependenztyp und eventuell die Wortartinformation benutzt. Das Kasusmerkmal hat nur indirekt zur Übersetzung beigetragen. Nach dem Parsing hat es seine Funktion erfüllt und braucht in der Baumstruktur nicht mehr wiedergegeben zu werden.

#### 3.5. Justierung des Regelsystems: Valenz

Tesnières bekanntester Begriff ist der der *Valenz*. Autoren, die vor Einsetzen der Tesnière-Rezeption arbeiten, stellen dabei die Begriffe *Dependenz* und *Valenz* meist nicht als ge-

trennt gegenüber. Solomon Kacnel'son (1948, 132) und A.W. de Groot (1949, 111) verstehen unter Valenz ganz allgemein die Fähigkeit des Wortes, syntaktische Verbindungen mit anderen Wörtern einzugehen. (Zur Bedeutung Kacnel'sons vgl. die Anmerkungen der Übersetzer in Kacnel'son 1987/1988, 1 Anm. 1.) Tesnière (1959/1982, 238) bezeichnet als Valenz die Fähigkeit eines Regens (er spricht vom Verb), eine bestimmte *Anzahl* abhängiger Aktanten zu regieren. In Tesnières Art, die Aktanten zu nummerieren, war eine Angabe dessen, was ich Dependenztyp nenne, wohl mitenthalten.

Für das Modell Metataxe ist es sinnvoll, eine präzisiertere Fassung des Valenzbegriffs zu verwenden. Ich definiere mit Engel (1977/1982, 96) Valenz als die für eine Unterart spezifische Fähigkeit, Dependenztypen zu regieren. Bei jeder Wortart, die überhaupt Dependenztypen haben kann, kann es Unterarten mit spezifischen Valenzeigenschaften geben. Die Valenz umfasst nicht nur die Zahl der Dependenztypen und nicht nur die Dependenztypen, sondern auch die syntaktische Form des Dependenztypen.

Engel leitet aus der Valenz die Unterscheidung zwischen Ergänzungen und Angaben her, wobei Ergänzungen valenzgebundene, also unterartenspezifische, und Angaben von allen Mitgliedern der Wortart gleichermaßen regierte Dependenztypen sind. Für die maschinelle Übersetzung ist die Unterscheidung zwischen Ergänzungen und Angaben nützlich, wobei manche theoretische Unklarheiten in der Verwendung, die die beiden Termini bei verschiedenen Autoren finden, durch zweckmäßige aber willkürliche Entscheidungen des Grammatikers beseitigt werden können. Zu diesen Unklarheiten gehört der Widerspruch zwischen einer bei vielen Autoren einschließlich Tesnière (1959/1982, 102) mitschwingenden intuitiven Festlegung der Ergänzungen als der „wichtigen“ Dependenztypen und der Tatsache, dass beispielsweise der von den allermeisten Autoren ohne Zögern als Ergänzung eingestufte Dependenztyp Subjekt Engels Definition im Deutschen nicht genügt, da jedes finite deutsche Verb ein Subjekt regiert und dieser Dependenztyp also gerade nicht auf eine Unterart beschränkt ist. Die intuitive Unterscheidung hat durchaus grammatischen Aussagewert, etwa wo es um die Regel geht, nach der ein bestimmtes Regens in ein und demselben Satz nur jeweils ein Exemplar eines Ergänzungstyps, aber ohne weiteres mehrere

Angaben gleichen Typs regieren kann. (Mehrere nebengeordnete Dependenztypen bilden zusammen *ein* Syntagma, so dass sie dieser Regel nicht widersprechen; vgl. 3.7) Da Information über die Anzahl möglicher Dependenztypen bei demselben Regens in der maschinellen Übersetzung für Bearbeitungsschritte wie das notorisch ambiguitätsgeplagte Parsing wichtig sind, nehme ich die Unterscheidung der Dependenztypen in Ergänzungen und Angaben in das Modell auf. Die geringe Zahl der Dependenztypen einer Sprache erlaubt notfalls eine willkürliche Zuordnung zu diesen beiden Gruppen, falls keine theoretisch besser untermauerte Definition zur Verfügung steht.

### 3.6. Feinjustierung des Regelsystems: Syntax und Wörterbuch

Die in Abb. 45.4 und 45.5 gezeigten Grundtypen syntaktischer Regeln konstituieren eine Basissyntax, die, wie in 3.3 beschrieben, der Justierung bedarf. Die beiden Grundregeltypen beschreiben jeweils nur einen Knoten und einen Ast beziehungsweise einen Ast und einen Knoten der Baumstruktur, wobei der Knoten durch Angabe der Wortart und der Ast durch einen Dependenztyp spezifiziert wird. Es gibt jedoch sprachliche Regelmäßigkeiten, die mit einer solchen Minimalspezifikation nicht zu erfassen sind. In 3.3 habe ich bereits am Beispiel des Objektkasus eine Regelmäßigkeit gezeigt, die zwei Knoten und den verbindenden Ast einbezieht und die überdies nicht damit auskommt, die Wortart des abhängigen Knotens festzulegen, sondern darüber hinaus ein bestimmtes Merkmal verlangt. Die Valenz (vgl. 3.5) ist eine Regelmäßigkeit, die zwar in vielen Fällen in Regeln des ersten Grundregeltyps formuliert werden kann, bei der jedoch die Angabe einer Wortart für den in dieser Regel vorkommenden Regensknoten nicht ausreicht. Regeln über valenzgebundene Dependenzrelationen machen es erforderlich, den Regensknoten auf eine Unterart einzuschränken.

Vor allem wenn es um die computerlinguistische Umsetzung solcher Regelmäßigkeiten geht, wird häufig erörtert, welche Information in die Syntax und welche in das Wörterbuch gehört. Das Modell Metataxe ist auch in dieser Frage von Zweckmäßigkeitsüberlegungen geleitet. Illustrationen wie Abb. 45.4 und 45.5 deuten bereits an, dass sich die syntaktischen Regeln des Modells in derselben Weise als Baumstrukturen darstellen lassen, wie die Strukturen der von ihnen be-

schriebenen Sätze. (Das Parsing wird dadurch zu einem Strukturabgleich zwischen Analyseobjekt und Regelmenge.) Auch die syntaktische Information, die man traditionellerweise in einem Wörterbuch sucht, lässt sich in Baumform darstellen. Das Modell Metataxe kennt daher keinen grundsätzlichen Unterschied zwischen Syntax und (syntaktischem) Wörterbuch. Wenn man in der computerlinguistischen Formalisierung Regelwerk und Wörterbuch getrennt halten möchte (etwa um in einem maschinellen Übersetzungssystem dem Anwender Zugriff auf das Wörterbuch, nicht aber auf die Syntax zu geben), dann lässt sich ein einfaches Trennungskriterium in den Baumstrukturen finden, die die Regeln darstellen. Baumstrukturen, deren inneres Regens ein konkretes Wort ist, werden dann dem Wörterbuch zugeordnet, während Regeln, deren inneres Regens eine Variable ist, in die Syntax gehören. Strukturen der ersten Art nenne ich *wortspezifische*, die der zweiten *konstruktionsspezifische Regeln*.

So klar diese Einteilung klingt, so willkürlich ist sie jedoch. Dies wird deutlich, wenn man das Problem redundanter Information einbezieht. Wenn zum Beispiel nur ein Teil der Mitglieder der Wortart Verb ein direktes Objekt regieren kann, dann muss irgendwo die Information notiert sein, um welche Verben es sich handelt. Notiert man im Wörterbuch bei jedem dieser Verben, dass sie ein direktes Objekt regieren können und welche Merkmale das Objekt trägt, so enthält das Wörterbuch bald eine Vielzahl gleichartiger, jeweils für ein anderes Wort spezifischer Regeln. Hier liegt es nahe, die redundante Information aus dem Wörterbuch herauszuziehen und jeweils nur einmal in Form einer konstruktionsspezifischen Regel zu formulieren, deren inneres Regens eine Variable für ein „Verb der Valenzklasse X“ ist. Die Wörterbucheinträge der betreffenden Verben enthalten dann nur ein Merkmal „Valenzklasse X“.

### 3.7. Komplexe Syntax

Arbeiten zur Syntax beschränken sich häufig auf die Regelmäßigkeiten des einfachen Haupt- und Nebensatzes. Auch die oben stehenden Abschnitte sind in dieser Weise geschrieben. Die einfachen Syntaxstrukturen, die durch die Grundregeltypen, die Elementarbäume sowie die besprochenen Justierungen in Form konstruktions- und wortspezifischer Regeln beschrieben werden, unterliegen jedoch zwei außerordentlich ambiguitäts-

trächtigen Erscheinungen, die schon vielen grammatischen Studien Stoff geboten haben: der Nebenordnung und der Ellipse.

Nebenordnungsstrukturen liegen dann vor, wenn zwei Dependenzien desselben unmittelbaren Regens durch eine (nebenordnende) Konjunktion verbunden sind. Von den möglichen Repräsentationen solcher Dependenzrelationen verwendet das Modell Metataxe eine Lösung, bei der die Konjunktion unmittelbares Regens der nebengeordneten Syntagmen ist. Hierdurch entsteht ein Syntagma üblicher Struktur mit einem einzigen inneren Regens (der Konjunktion) und zwei Dependenzien. Der Dependenztyp, mit dem das nebengeordnete komplexe Syntagma von einem äußeren Regens abhängt, verdoppelt sich so, dass die beiden abhängigen Syntagmen von der Konjunktion mit demselben Dependenztyp abhängen. In den Baumstrukturen kennzeichne ich diese verdoppelten Dependenztypen mit einem Etikettsuffix (auf Deutsch -N „Nebenordnung“). Die Nebenordnung ist binär, kann aber rekursiv angewendet werden. Abb. 45.5 zeigt (mit Tesnières Beispielsatz) zwei nebengeordnete Subjekte, Abb. 45.6 eine Dreieraufzählung. Das Komma, das hier ähnlich wie eine Konjunktion fungiert, wird wie ein Wort behandelt und steht auf dem Knoten der Baumstruktur.

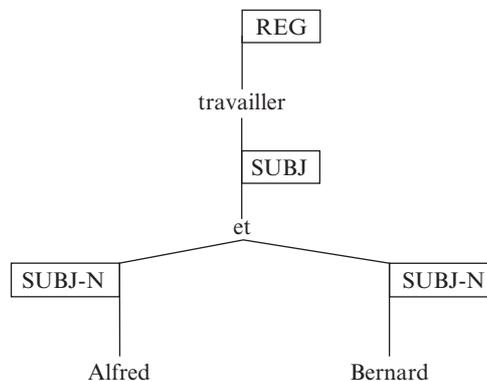


Abb. 45.5: Zwei nebengeordnete Subjekte

Nebenordnungsstrukturen können auch auf verschiedene andere Weisen repräsentiert werden. Ich habe die wichtigsten Alternativen diskutiert und die Entscheidung für die hier gezeigte Lösung begründet (Schubert 1987, 104–119). Ein wesentlicher Vorteil dieser Lösung liegt darin, dass die in der einfachen Syntax gültigen Regelmäßigkeiten über die Anzahl von Dependenzien (vgl. 3.5) bei dieser Lösung auch für Nebenordnungsstrukturen

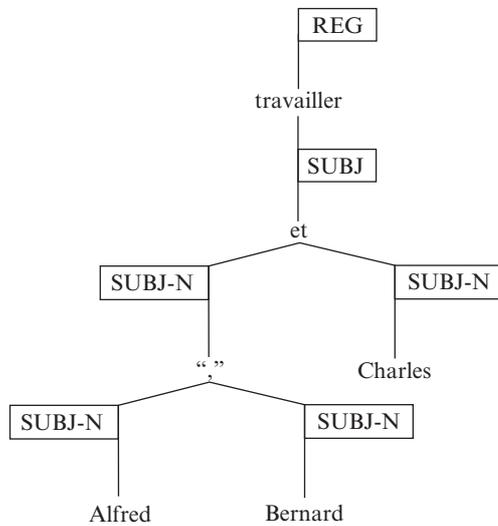


Abb. 45.6: Drei nebengeordnete Subjekte

tionen gelten, da diese nach außen hin als ein einziges Syntagma mit nur einem inneren Regens auftreten. Die Konjunktion erhält hierbei die Funktion des „Weiterkopierens“ syntaktischer Regeln und Merkmale in aufsteigender und absteigender Richtung im Baum, ähnlich wie dies von den Vererbungsregeln anderer Grammatikmodelle bekannt ist.

Die Beschreibung in den oben stehenden Absätzen geht der Einfachheit halber von dem sehr häufigen Fall aus, dass das komplexe Nebenordnungssyntagma als Ganzes von einem äußeren Regens abhängt, wie in Abb. 45.5 und 45.6 gezeigt. Darüber hinaus gibt es jedoch auch nicht selten Sätze, in denen die inneren Regentien zweier Hauptsätze nebengeordnet sind. Das Modell Metataxe repräsentiert diese Konstruktion analog zu den gezeigten Beispielen, wobei die Konjunktion dann Hauptregens des komplexen Satzes ist und der Dependenztyp, der durch die Konjunktion verdoppelt wird, der leere Dependenztyp Regens (REG) ist. Mit diesem Etikett markiere ich das Hauptregens eines Satzes. Das Etikett bekommt erst dann eine Funktion, wenn satzübergreifende Syntax betrieben wird.

Die hier gegebene Beschreibung geht davon aus, dass die Nebenordnung zwei Dependenztypen gleichen Dependenztyps zusammenstellt. Eine solche Regel ist für das Parsing hilfreich. Sie begrenzt jedoch die Gestaltungsfreiheit bei der Festlegung der Dependenztypen (vgl. 3.2). Insbesondere im Bereich

der freien Angaben sind in manchen Sprachen recht unterschiedliche Konstruktionen nebenordnungsfähig. Ordnet man diesen verschiedene Dependenztypen zu, so müssen Regeln über miteinander nebenordnungsfähige und nicht nebenordnungsfähige Dependenztypen aufgestellt werden. Geht man umgekehrt vor und ordnet alles, was nebengeordnet werden kann, demselben Dependenztyp zu, so erhält man einige recht heterogene Dependenztypen, die jeweils sehr häufig vorkommen. Welche Lösung man bevorzugt, hängt davon ab, wie zweckmäßig es ist, in der Strukturübertragung mit vielen beziehungsweise mit wenigen Dependenztypen zu arbeiten.

Dass die Nebenordnung zu den notorischen Problemfällen der Syntax und ihrer computerlinguistischen Formalisierung gehört, hat vor allem zwei Ursachen. Eine davon ist die Ellipse, von der unten die Rede ist. Die andere ist die Tatsache, dass Nebenordnung auf allen Ebenen der Syntax möglich ist, von unvollständigen Wörtern (*Im- und Export*) bis hin zu ganzen Satzgefügen und Textteilen. Hierdurch wird das Erkennen der gemeinten Dependenzrelation oft sehr schwierig. Die Syntax hat hier wie in anderen Fällen jedoch nur die Aufgabe, den Entscheidungsraum zu beschreiben. Zum Fällen der im Analysevorgang notwendigen Entscheidungen fehlen ihr Mechanismen und Kriterien (vgl. 2.2).

Ein sprachliches Phänomen, das ebenfalls der Beschreibung des Entscheidungsraums erhebliche Schwierigkeiten bereitet, ist die Ellipse. Die Benennung dieses Phänomens deutet auf die Vorstellung, dass etwas weggelassen sei. Eine solche Vorstellung bringt die empirisch vorgefundene sprachliche Wirklichkeit in einen Bezug zu einem Konstrukt wie etwa dem eines vollständigen, nichtelliptischen Satzes. Ob überhaupt theoretische Konstrukte dieser Art verwendet werden sollen und wie sie im Einzelnen zu gestalten sind, liegt dabei in der Hand des Grammatikers. Den meisten Dependenzmodellen liegt der Rückgriff auf Konstrukte sehr viel ferner als den Modellen der generativen Tradition, an deren Beginn der Transformationsgedanke stand. Im Modell Metataxe ist es sinnvoll, weitgehend ohne Ellipsen auszukommen, solange von einfachen Haupt- und Nebensätzen die Rede ist. Was in diesem Bereich bisweilen als Ellipse erklärt wird, findet seinen Niederschlag im Modell Metataxe eher in der Form eines Verzichts auf allzu kategori-

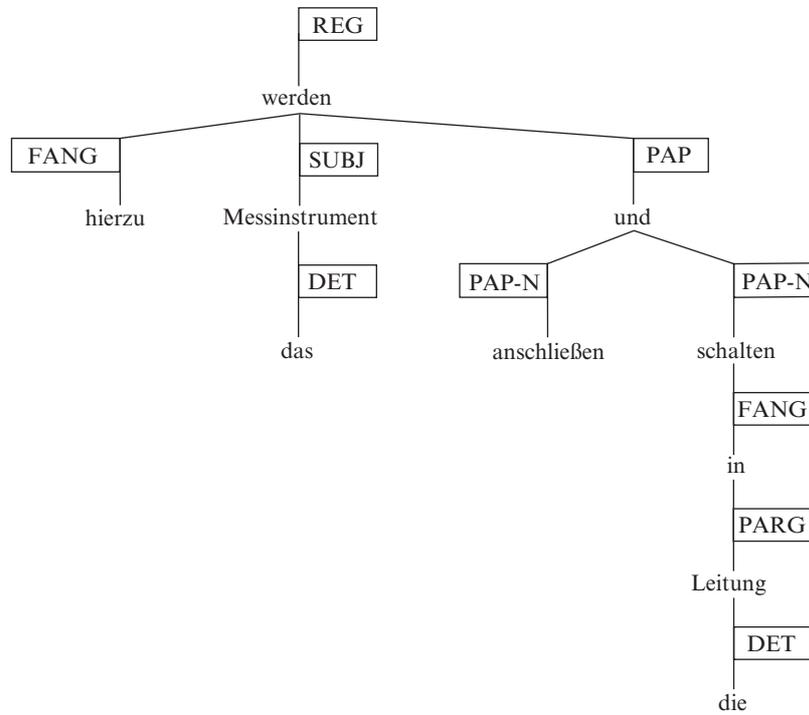


Abb. 45.7: Nichtelliptische Konstruktion: nebengeordnete Passivpartizipien

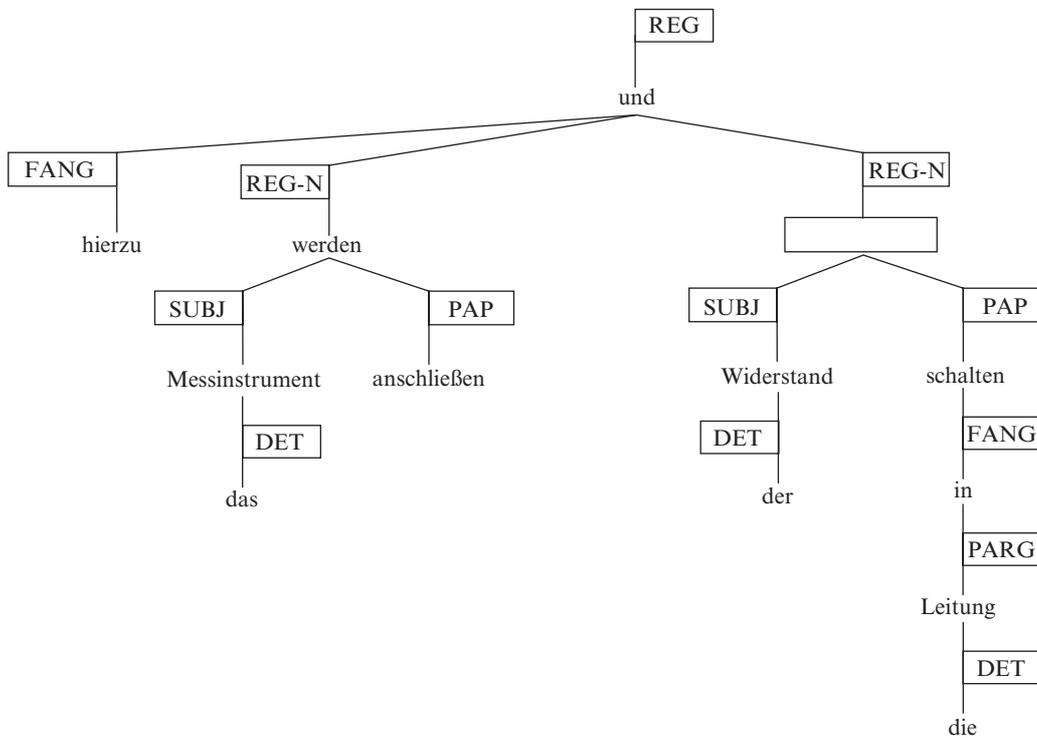


Abb. 45.8: Elliptische Konstruktion mit leerem Knoten

sche Einstufung bestimmter Dependenzien als obligatorisch. Hierdurch verliert auch die Unterscheidung von Ergänzungen und Angaben einen Teil der Bedeutung, die sie in anderen Modellen hat.

Das Modell Metataxe macht jedoch Gebrauch von der Vorstellung der Ellipse, wo es sich um Nebenordnungen auf der Ebene des Haupt- oder Nebensatzes handelt. In solchen Konstruktionen lässt eine große Zahl von Sprachen im ersten wie im zweiten Glied der Nebenordnung spezifische Verhältnisse zu, die sich am konsistentesten erklären lassen, wenn man annimmt, dass bestimmte Dependenzien von einem Regens abhängen, das nicht dasteht. Die Regenseigenschaften dieses als Konstrukt eingefügten leeren Wortes lassen sich mit recht guter Regelmäßigkeit von denen seines Nebenordnungsgegenstücks ableiten. Der Satz *Hierzu wird das Messinstrument angeschlossen und in die Leitung geschaltet* erfordert noch keine Erklärung durch Ellipse. Hier regiert das finite Verb *werden* das Subjekt (Abb. 45.7). Dagegen weist der Satz *Hierzu wird das Messinstrument angeschlossen und der Widerstand in die Leitung geschaltet* eine Nebenordnungsellipse auf, bei der ein leerer Knoten mit den Regenseigenschaften des ihm nebengeordneten Knotens *werden* angesetzt wird (Abb. 45.8).

In den fachsprachlichen Textsorten, die Haupteinsatzgebiet der maschinellen Übersetzung sind, ist die Nebenordnungsellipse die weitaus häufigste. Andere, vor allem umgangssprachliche Ellipsen, die in der linguistischen Literatur oft erörtert werden, spielen hier keine herausragende Rolle.

### 3.8. Textsyntax

Syntax wird oft im Wesentlichen als Lehre vom Satzbau betrieben. Spätestens beim Arbeitsgang der Übertragung zeigt sich jedoch, dass ein Teil der zum Übersetzen eines Satzes erforderlichen Information in dem zu übersetzenden Text, jedoch außerhalb des betreffenden Satzes gesucht werden muss (vgl. 4.3). Eine Übersetzungsgrammatik ist daher ohne eine textgrammatische Komponente unvollständig.

Eine Syntax auf Textebene hat die Regelmäßigkeiten der Kombinierbarkeit der Elemente eines Textes zu beschreiben. Solange es sich nicht als sinnvoll erweist, Zwischenebenen anzusetzen, geht es hierbei also vor allem um Regeln für die Kombination von Sätzen. Bei Versuchen, derartige Regeln zu beobachten und in grammatisch konsistente Form zu

bringen, fällt auf, dass die Kombinierbarkeit der Elemente in sehr hohem Maße der Wahlfreiheit des Sprechers beziehungsweise Autors unterliegt und somit inhaltlicher Art ist. Wo der Sprecher Wahlfreiheit hat, bietet das System Ansatzpunkte für Einflüsse des kommunikativen Gebrauchs der Sprache in der Sprachgemeinschaft. Es wird deutlich, dass auf Textebene auf der Formseite des sprachlichen Zeichens nur einige wenige Regelmäßigkeiten operieren. Diese haben Einfluss auf die Textstruktur, determinieren sie jedoch keineswegs in einem Umfang, wie dies bei der Struktur von Sätzen und Syntagmen der Fall ist. Die Balance zwischen textsyntaktischem Eröffnen des Entscheidungsraums und Entscheidungsmechanismen semantischer, pragmatischer und außersprachlicher Art ist hier gegenüber der Satzebene stark in Richtung auf die inhaltsbezogenen Mechanismen verschoben.

In einer übersetzungsorientierten Studie haben Bart Papegaaïj und ich drei zentrale Mechanismen der Textkohärenz beschrieben: die Kohärenz der Referenz, die Kohärenz der kommunikativen Gliederung und die Kohärenz der Ereignisse („coherence of entities“, „coherence of focus“, „coherence of events“: Papegaaïj/Schubert 1988, 159–199). Bei der Kohärenz der Referenz geht es um Aussagen in einem Text, die auf dieselben außersprachlichen Objekte und Ereignisse verweisen. Sie hat einen syntaktischen Reflex in Sprachen, in denen eine Bedeutungskomponente wie Vorerwähntheit oder „shared knowledge“ einen grammatikalisierten Ausdruck findet, etwa in der Bestimmtheit des Substantivs. Ein weiterer syntaktischer Reflex liegt in der syntaktischen Beschränkung der Auswahl möglicher Korrelate anaphorischer und kataphorischer Verweise durch Pronomen und andere deiktische Mittel (einen dependenzuellen Lösungsansatz zeigen Lappin/McCord 1990). Bei der Kohärenz der kommunikativen Gliederung geht es unter anderem um die Thema-Rhema-Gliederung von Sätzen. Ihr syntaktischer Reflex liegt je nach Sprache in der Bestimmtheit des Substantivs oder beispielsweise in der Syntagmenfolge mit den für die jeweilige Sprache spezifischen Mechanismen der Syntagmenumstellung (Wortfolgeänderung, Diathesewechsel, emphatische Konstruktion usw.). Die Kohärenz der Ereignisse betrifft die Aufeinanderfolge außersprachlicher Ereignisse. Sie findet ihren Niederschlag vor allem in bestimmten Merkmalen der präzifizierenden Wörter und Wortgruppen des

Textes. Ihr syntaktischer Reflex sind unter anderem die Wahl und die Abfolge der Tempus-, Modus- und Aspektformen des Verbs.

Die textgrammatischen Kohärenzphänomene sind noch lange nicht so gründlich untersucht, wie dies für eine computerlinguistische Umsetzung wünschenswert wäre. Sie sind daher Gegenstand unterschiedlichster Forschungstätigkeit auf der linguistischen wie auf der computerlinguistischen Seite. Das Modell Metataxe kann am Ende des Übersetzungsvorgangs, in der Synthesephase, recht deutlich aufzeigen, welche Information auf Textebene für eine vollständige Übersetzung notwendig ist (vgl. 4.3).

#### 4. Die eigentliche Metataxe: dependenzsyntaktische Strukturübertragung

Maschinelle Übersetzungssysteme sind heute durchweg nach der Transferarchitektur aufgebaut (vgl. 2.1), bei der nacheinander ein einsprachiger Analysevorgang, ein zweisprachiger Übertragungsvorgang und ein einsprachiger Synthesevorgang durchlaufen werden. Die beiden einsprachigen Module lassen sich wiederverwenden: Beispielsweise kann für die Übersetzungsrichtung Deutsch-Französisch und die Übersetzungsrichtung Deutsch-Russisch dasselbe deutsche Analysemodul eingesetzt werden – was allerdings eine strikte, modulare Trennung der horizontalen Entwurfsebenen Linguistik, Formalisierung und Implementierung (vgl. 1) sowie der vertikalen Bearbeitungsschritte Analyse, Übertragung und Synthese voraussetzt. Strebt man eine solche Wiederverwendung an, so wird man bestrebt sein, einen möglichst großen Anteil der Gesamtbearbeitung in die einsprachigen Module zu verlegen und den Übertragungsvorgang so klein wie möglich zu halten.

Es ist daher nicht verwunderlich, wenn Arbeiten über das Grammatikmodell Metataxe sich sehr viel ausführlicher mit einsprachiger Syntax als mit der eigentlichen Metataxe zu beschäftigen haben. Dieser Effekt wird verstärkt durch die Tatsache, dass der Schritt von der unbearbeiteten Sprache zu einer zweckmäßigen künstlichen Struktur wesentlich aufwendiger ist als nachfolgende Umformungsschritte, die mit diesen künstlichen Strukturen arbeiten. Aus diesem Grunde ist die Analysesyntax ungleich umfangreicher als die Synthese.

*Metataxe* ist die Strukturumwandlung, die ein Satz beim Übersetzen durchläuft. Während Tesnière (1959/1982, 283) vor allem Abweichungen von einer stillschweigend als im Wesentlichen gleich angenommenen gemeinsamen syntaktischen Struktur von Ausgangs- und Zielsprache ins Auge fasst, gehe ich im Modell Metataxe davon aus, dass die Struktur jedes Satzes im Übertragungsschritt vollständig zerlegt und in der Zielsprache neu wieder aufgebaut wird. Durch diese Vorgabe sind die Syntaxen der Ausgangs- und Zielsprache autonom und es spielt keine Rolle, ob Wortarten, Merkmale und Dependenztypen beider Sprachen ganz, teilweise oder gar nicht deckungsgleich sind.

Die Komplexität des Übertragungsvorgangs liegt wesentlich darin begründet, dass hier zwei Bearbeitungsstränge nebeneinander durchlaufen werden, die jeweils für sich die Abfolge eines entscheidungsraumeröffnenden Mechanismus und eines Entscheidungsmechanismus aufweisen und die überdies ineinander verschachtelt sind. Dies sind die Mechanismen der eigentlichen Metataxe oder Strukturübertragung einerseits und die lexikalischen Übertragung andererseits. Bei der lexikalischen Übertragung wird der Entscheidungsraum durch das Wörterbuch eröffnet, das für ein ausgangssprachliches Wort oder Syntagma oft mehrere alternative Übersetzungsentsprechungen zur Verfügung stellt. Der Entscheidungsmechanismus hat dann die Aufgabe, anhand inhaltlichen Wissens über den Kontext und die Übersetzungsentsprechungen die beste Lösung auszuwählen. In der eigentlichen Metataxe ist mit einem sehr viel begrenzteren Entscheidungsraum zu rechnen als bei der Analyse, solange das System so konstruiert ist, dass die Analyseambiguität vor Beginn der Übertragung gelöst wird. Legt man dagegen den Systementwurf so an, dass ein Teil der Entscheidungsaufgaben aus der Analyse ungelöst in die Übertragung transportiert wird, dann laufen mehrere konkurrierende Metataxevorgänge ab, aus deren Ergebnissen ein neuer Entscheidungsraum entsteht.

Die eigentliche Metataxe habe ich auf der Ebene des linguistischen Systementwurfs mehrfach unter verschiedenen Aspekten beschrieben (Schubert 1986, 166–191; 1987, 130–194; 1988b; 1989c). Eine operative Lösung auf der Ebene der computerlinguistischen Formalisierung beschreibt Job van Zuijlen (1989). Nach diesem Modell ausgearbeitete Metataxen liegen für die Sprachen-

paare Englisch-Esperanto (Maxwell 1989), Esperanto-Französisch (Tamis 1989) und Isländisch-Esperanto (Briem, unveröffentlicht) vor.

#### 4.1. Metataxe: Regelwerk und Wörterbuch

Die eigentliche Metataxe hat die Aufgabe, dependenzsyntaktische Baumstrukturen der Ausgangssprache in ebensolche Baumstrukturen der Zielsprache umzuwandeln. In erster Annäherung kann man sagen, dass dabei die Knoten der Bäume von der lexikalischen Übertragung und die Äste von der Metataxe übertragen werden.

Ein erster Ausgangspunkt für die Metataxe ist die Aufgabe, jede syntaktisch zulässige Struktur der Ausgangssprache im Wege einer vollständigen Abbildung (vgl. 2.4) in eine übersetzungsäquivalente, syntaktisch zulässige Struktur der Zielsprache umzusetzen. Ausgangsmaterial sind Baumstrukturen der Ausgangssprache, deren Knoten Wörter mit Merkmalen und deren Äste Dependenztypen-etiketten tragen.

Ein zweiter Ausgangspunkt ist das Bestreben, wie bei der einsprachigen Syntax (vgl. 3.3) auch bei der Metataxe mit möglichst einfachen, möglichst allgemeinen und möglichst häufig anwendbaren Regeln zu arbeiten und das so aufgebaute Regelwerk nur dort durch komplexere Regeln zu justieren, wo die sprachliche Realität dies erfordert.

Als dritter Ausgangspunkt soll die Vorgabe gelten, dass die Metataxe lieber falsch als gar nicht übersetzen soll. Dies ist eine praktische Bedingung. Sie sagt aus, dass es zweckmäßig ist, dafür zu sorgen, dass das Regelwerk der Metataxe für jede denkbare syntaktische Struktur der Ausgangssprache eine Übersetzungsentsprechung in der Zielsprache erzeugen kann, und dass es im Falle unvollständiger oder unzureichender Regeln zweckdienlicher ist, wenn als Übertragung doch eine syntaktisch korrekte, wenn auch vielleicht nicht im wünschenswerten Maße übersetzungsäquivalente Struktur geliefert wird, als dass der Übersetzungsvorgang ganz anhält oder Teile des Eingabetextes auslässt.

Die in diesen drei Ausgangspunkten formulierten Vorgaben der Vollständigkeit und der modularen Einfachheit führen zu der Überlegung, dass zunächst für alle Elemente der ausgangssprachlichen Baumstrukturen isolierte Metataxeregeln aufgestellt werden sollten. Diese Elemente sind die Wortarten und die Dependenztypen. Übertragungsregeln für die Wortarten liefert die lexikale

Übertragung, indem sie als Übersetzungsentsprechungen zielsprachliche Wörter anbietet, die jeweils einer Wortart angehören. Der Metataxe bleibt daher die Aufgabe, die Dependenztypen zu übertragen. Die grundlegende Ebene der Metataxeregeln bildet daher eine Gruppe von Regeln, die lediglich einen Ast des Baums mit seinem Etikett übertragen. Für jeden Dependenztyp der Ausgangssprache muss es eine solche Regel geben. Eine solche Regel überträgt beispielsweise den deutschen Dependenztyp Subjekt in den englischen Dependenztyp Subjekt. In der grafischen Darstellung tragen die Dependenztypenetiketten jetzt Sprachpräfixe (Abb. 45.9).

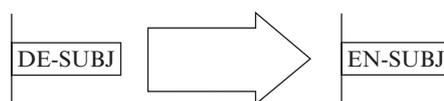


Abb. 45.9: Einfache Metataxeregels

Regeln des in Abb. 45.9 gezeigten Typs sorgen (zusammen mit einem Mechanismus der lexikalischen Übertragung für das Vorgehen bei „unbekannten“, d. h. im Wörterbuch fehlenden Wörtern) dafür, dass jede von der Analyse erzeugte Baumstruktur eine zielsprachliche Entsprechung bekommt. Enthielte das Regelwerk jedoch nur diese allgemeinsten Regeln, wäre das Übertragungsergebnis sehr oft falsch. Es bedarf also einer gründlichen Justierung durch komplexere Regeln.

Wie bei den Regeln der einsprachigen Syntax lassen sich auch in der Metataxe Regeln unterschiedlicher Spezifik unterscheiden. Insbesondere sind konstruktionsspezifische und wortspezifische Regeln zu unterscheiden, wobei wortspezifische ins zweisprachige Wörterbuch gehören. Hier liegt also einer der Punkte, an denen Metataxe und lexikale Übertragung verflochten sind.

Abb. 45.10 und 45.11 zeigen zwei Metataxeregeln für die englisch-deutschen Entsprechungen (*someone*) *is given* (*something*) → (*jemandem*) *wird* (*etwas*) *gegeben* und (*someone*) *is given* (*something*) → (*jemand*) *erhält* (*etwas*). Es handelt sich um eine wortspezifische Metataxeregels mit dem inneren Regens *be*, das eine lexikale Alternative (*werden*, *erhalten*), konstruktionsspezifische Strukturumwandlungen (englisches Subjekt wird deutsches indirektes Objekt) und wortspezifische Strukturumwandlungen (englisches *be* mit Dependens *give* mit Merkmal Partizip Perfekt wird deutsches *erhalten* mit Merkmal

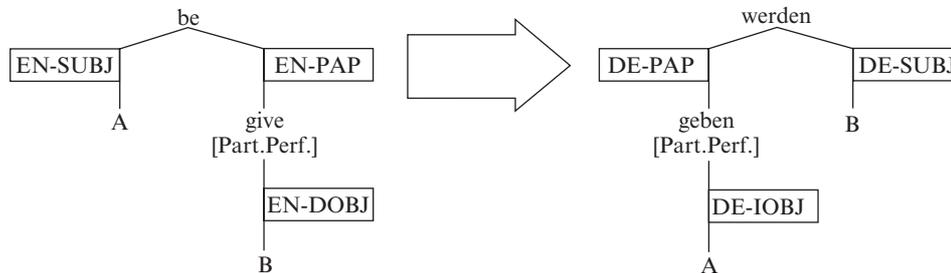


Abb. 45.10: Komplexe Metataxaregel

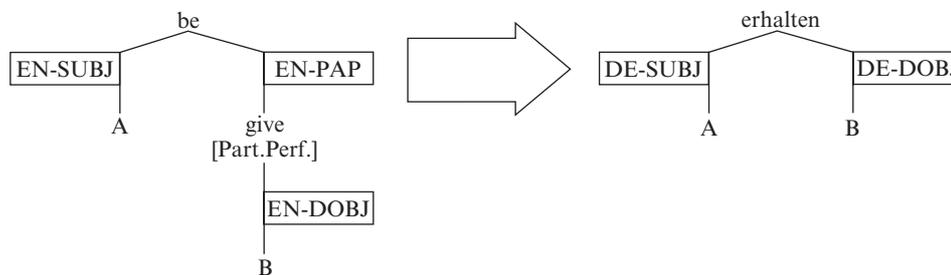


Abb. 45.11: Komplexe Metataxaregel

Aktiv) zeigt. Beispiele wie diese sind oft Anlass lebhafter Diskussion, die sich zum Teil auf die einsprachige Syntax bezieht. So ist es auf den ersten Blick wenig intuitiv, wenn das Passivhilfsverb inneres Regens eines Syntagmas wie des hier gezeigten ist, sodass die betreffende Regel eine von Hunderten ähnlicher Regeln über das Verb *be* wird. So verfahren Iordanskaja (1964, 240) und Tarvainen (1985, 60) bei verschiedenen Hilfsverbgefügen in gleicher Weise, während David Hays (1963, 190, 198; 1964, 521) das Partizip zum Regens des finiten Verbs macht. In 5 komme ich hierauf zurück.

#### 4.2. Der Metataxevorgang

Ein vollständiges Metataxeregelwerk besteht aus einer größeren Zahl allgemeinerer und Tausenden wortspezifischer Regeln – weil jeder Eintrag im zweisprachigen Wörterbuch eine Metataxaregel ist (vgl. 4.1, Abb. 45.10 und 45.11). Auf Besonderheiten einzelner Sprachenpaare soll hier nicht eingegangen werden. Von übergreifendem Belang ist jedoch die Frage, wie im Metataxevorgang aus dieser großen Regelmenge die richtige Auswahl getroffen werden soll. In 4.1 war bereits andeutungsweise von verschiedenen Ebenen von Regeln die Rede, deren unterste die ein-

fachsten (oder merkmallösen, vgl. Schubert 1987, 147) Metataxeregel bilden. Beide Fragen stehen miteinander in Zusammenhang.

Ebenso wie beim Aufbau des einsprachigen und des zweisprachigen Regelwerks ist es auch bei der Gestaltung des Bearbeitungsvorgangs zweckmäßig, mit möglichst kleinen, möglichst häufig einsetzbaren Bearbeitungsregeln zu operieren. Insbesondere ist festzulegen, welche Teilstrukturen einer ausgangssprachlichen Baumstruktur jeweils von einer Bearbeitungsregel erfasst wird und in welcher Reihenfolge die Teilstrukturen abgearbeitet werden.

In Mathematik und Informatik gibt es eine Reihe gut untersuchter Abarbeitungspfade durch Graphen und Baumstrukturen. Für das Modell Metataxe erscheint es mir jedoch sinnvoll, eine Lösung zu wählen, die im Wesentlichen sprachlichen Kriterien folgt und mit möglichst wenig außersprachlicher Steuerung auskommt. Wenn ein ausgangssprachlicher Baum schrittweise gelesen und durch Anwendung von Regeln der Metataxe und der lexikalen Übertragung ein zielsprachlicher Baum aufgebaut wird, dann könnte ein umfangreicher Mechanismus erforderlich werden, der protokolliert, was auf der ausgangssprachlichen Seite bereits bearbeitet ist

und wo neue Bauteile des zielsprachlichen Baums anzufügen sind. All dies vermeidet das Modell Metataxe durch die Idee der *hybriden Baumstruktur*. Nach dieser Vorgabe entsteht der zielsprachliche Baum, indem die Komponenten des ausgangssprachlichen Baums Stück für Stück durch Komponenten eines übersetzungsäquivalenten zielsprachlichen Baums ersetzt werden. Der Metataxevorgang arbeitet also in einer einzigen Baumstruktur, die zunächst nur Komponenten der Ausgangssprache enthält und im Laufe der Bearbeitung nach und nach zu einer Baumstruktur der Zielsprache wird. Bevor dieses Ziel erreicht ist, handelt es sich daher um einen hybriden Baum mit Komponenten aus beiden Sprachen.

Der Metataxevorgang verläuft dann in folgenden Schritten, auf die ich unten im Einzelnen weiter eingehe:

- (1) Die Bearbeitung beginnt beim Hauptregens der Übersetzungseinheit.
- (2) Aus dem Gesamtmetataxeregelwerk wird die Menge der auf das Hauptregens *anwendbaren* Metataxeregeln ermittelt.
- (3) Aus der Menge der *anwendbaren* Metataxeregeln wird die Regel mit der höchsten *Regelpriorität* angewendet. Der Baum wird hierdurch zu einer hybriden Baumstruktur.

Es entsteht eine Grenze zwischen zielsprachlichen und ausgangssprachlichen Komponenten. (Komponenten sind Wörter mit Merkmalen sowie Dependenztypen.) Die Grenze bildet immer eine kontinuierliche Linie, über der nur zielsprachliche und unter der nur ausgangssprachliche Komponenten stehen. Der folgende Schritt ist ein Element willkürlicher, nicht sprachlich motivierter Steuerung:

- (4) Aus der Menge der direkt unterhalb der Grenze stehenden ausgangssprachlichen Komponenten wird die am weitesten links stehende zum Gegenstand des nächsten Bearbeitungsschrittes.
- (5) Die Bearbeitung beginnt wieder bei Schritt (2).
- (6) Die Bearbeitung endet, wenn alle Komponenten in die Zielsprache übertragen sind.

*Anwendbar* ist eine Regel, deren ausgangssprachliche Seite mit einem Teil des zu übertragenden Baumes deckungsgleich ist, dessen inneres Regens die in Schritt (4) identifizierte Komponente ist. Regeln können Wörter oder

Variablen für Wörter, Merkmale oder Variable für Merkmale und Dependenztypen oder Variablen für Dependenztypen enthalten. Von zwei anwendbaren Regeln genießt diejenige *Priorität*, die komplexer ist, wobei Komplexität an der Anzahl der Komponenten der ausgangssprachlichen Seite der Regel gemessen wird. Erst wenn zwei Regeln durch all diese Vergleiche nicht in eine Prioritätsreihenfolge gebracht werden können, wird ein willkürliches Kriterium zugelassen. Die Dependenzsyntax ist so gestaltet, dass die Anordnung mehrerer Dependenztypen desselben Regens von links nach rechts keine Rolle spielt. Sie bleibt daher auch beim Abgleich zwischen Regel und Baum unberücksichtigt. Die hier beschriebene Abgleichstechnik entspricht in ihrer Funktion im Wesentlichen den Grammatikformalismen, die Bedingungen über Attribute und ihre Werte formulieren (z. B. Attributgrammatiken).

#### 4.3. Synthese und Textgrammatik

Ergebnis des Strukturübertragungsarbeitsgangs ist eine Baumstruktur in der Zielsprache. Ihre Äste tragen zielsprachliche Dependenztypenetiketten. Die Knoten tragen Wörter der Zielsprache. Die Wörter stehen in der Grundform und sind mit Merkmalen versehen. Hierbei handelt es sich zunächst nur um die direkt übersetzbaren Merkmale (vgl. 3.4).

Der Syntheseschritt hat die Aufgabe, die durch die von der Grammatik (der Analyse- und der Übertragungssyntax) gesteuerten Mechanismen hinzugefügten metasprachlichen Elemente dieser Baumstruktur nach den Regeln der Zielsprache in sprachliche Elemente umzuformen. Im Wesentlichen geht es hierbei um drei Typen von Elementen: (1) die Dependenztypenetiketten werden in Merkmale der nicht direkt übersetzbaren Art umgewandelt, (2) die Merkmale (der direkt und der nicht direkt übersetzbaren Art) werden in sprachliche Form umgewandelt, was seinen Ausdruck in Flexionsformen, Funktionswörtern oder Wortfolgeanordnungen finden kann und (3) der zweidimensionale Baum wird linearisiert, was bedeutet, dass die Wörter in einer eindimensionalen Kette angeordnet werden. Welches Merkmal welchen Ausdruck findet, ist in hohem Maße einzel-sprachspezifisch.

In den meisten Fällen reicht der Synthesevorgang mit diesen drei Elementumwandlungen nicht aus, um den zielsprachlichen Text vollständig zu spezifizieren. Es bleiben Entscheidungsmöglichkeiten offen. Je nach Spra-

chenpaar kann es sich hierbei um Alternativen in der Wortfolge (genauer: der Syntagmafolge) oder der Wahl bestimmter Merkmale wie z. B. der Bestimmtheit von Substantiven handeln. Wenn im Syntheseschritt solche Entscheidungsalternativen offen bleiben, so bedienen sich die meisten operativen maschinellen Übersetzungssysteme willkürlicher Mechanismen, indem sie beispielsweise eine Normwortfolge einhalten oder in der Frage der Bestimmtheit die Ausgangssprache kopieren. Für die Grammatik sind diese offenen Entscheidungen jedoch ein Hinweis darauf, dass entscheidungsunterstützendes Wissen fehlt. Wenn die Metataxe und die ihr vorausgehende Analyse auf Satzebene operieren, so ist das fehlende Wissen oft textlinguistischer Art. Hierher gehören Regelmäßigkeiten der Thema-Rhema-Gliederung, der Bildung von Ereignisketten, der Anaphern- und Kataphernbildung usw. Gelänge es, dieses Wissen in geeigneter Form greifbar zu machen und zu formalisieren, würde ein wichtiger Schritt zu *kohärenten Texten* als Ergebnis der maschinellen Übersetzung möglich (vgl. Papegaaij/Schubert 1988, 159–199).

Wie Einzelschritte des Synthesevorgangs verlaufen und wo die Ansatzpunkte für textlinguistische Entscheidungssteuerung liegen, habe ich in knapper Form skizziert (Schubert 1992b).

## 5. Modellentwurf und dependenztheoretisches Umfeld

Das Modell Metataxe ist auf die computerlinguistische Praxis gerichtet und leistet gleichzeitig einen Beitrag zur laufenden Diskussion um die Theoriebildung in einem Bereich der Grammatik. Ein Blick auf einige zentrale Entwurfsentscheidungen soll daher die Darstellung des Modells abrunden und den Vergleich mit konzeptionell verwandten Theorien erleichtern.

Im Bereich der eigentlichen Metataxe liegt außer Tesnières ursprünglicher Darstellung wenig Vergleichbares vor. Im Bereich der einsprachigen Syntax reiht sich das Modell Metataxe jedoch in eine lange Kette dependenzorientierter Theorien und Grammatiken ein. Grundlegende Entwurfsalternativen liegen hier vor allem im Bereich der grafischen Darstellung der syntaktischen Struktur, wobei insbesondere die Fragen der Wiedergabe von Inhalts- und Funktionswörtern, der Wahl zwischen Graph und Baum, der Projek-

tivität sowie der leeren und der abstrakten Knoten eine Rolle spielen. Weiter ist die Frage des Verhältnisses zwischen Syntax und Semantik mit den sich hieraus ergebenden Konsequenzen für die syntaktische Struktur interessant.

Grundlage eines Grammatikmodells ist die Wahl der Elemente und der Relationen, mit denen das Modell das sprachliche System beschreiben soll. Das Wort ist für die allermeisten Grammatiker das naheliegendste Element, wenn auch durchaus Lösungen sinnvoll sein können, die beispielsweise auf dem Morphem aufbauen. Das Modell Metataxe untersucht die Konkmitanz der Elemente des Systems distributionell und leitet hieraus eine Klassifikation der Elemente und eine Klassifikation der Relationen ab. Es gelangt so zu Wortarten und Dependenztypen.

Die meisten der nach diesem Modell ausgearbeiteten Dependenzsyntaxen sind sehr genau darin, jedes einzelne Wort des Textes auch in der Baumstruktur zu platzieren. (Dolinina 1969, 303 zeigt anschaulich, wie sich Tesnières Darstellung hier von späteren Dependenzgrammatiken unterscheidet.) Spätestens bei der Kopplung syntaktischer Strukturen an semantische Entscheidungsmechanismen stellt sich jedoch die Frage, ob es empfehlenswert ist, Inhalts- und Funktionswörter durchgängig gleich zu behandeln. Sadler (1989, 44) beschreibt einen semantischen Mechanismus, bei dem die Gesamtaussage eines Satzes (oder auch einer kleineren oder größeren Texteinheit) in eine Reihe von Elementaraussagen der Form X R Y zerlegt wird, wobei X und Y für Inhaltswörter stehen und R ein semantischer Relator ist. Soweit Sadlers semantische Relationen syntaktischen entsprechen, ist X direktes oder indirektes Regens von Y. Als Relatoren verwendet Sadler teils Funktionswörter wie Präpositionen, teils Kennzeichen der zentraleren syntaktischen Relationen wie der Subjekt-, der Objekt- oder der Attributrelation. Wollte man die einsprachige Dependenzsyntax so gestalten, dass sie einem Semantikmodell wie dem Sadlers von vornherein entgegenkommt, so wäre zu überlegen, ob ein Teil der Funktionswörter als Indizien für Merkmale eingestuft werden sollten. Dies hätte zur Folge, dass sie beim Parsing in ihrer Indizfunktion erkannt werden könnten und nach Notierung des so angezeigten Merkmals nicht weiter als Wort repräsentiert zu werden brauchten. In manchen Fällen bedeutet eine solche Lösung nur eine geringe Abweichung von der oben dargestell-

ten Form der Dependenzsyntax, in anderen kommt es zu einer eingreifenden Umgestaltung des Modells. Ein Beispiel für eine geringfügige Veränderung wäre die Umsetzung bestimmter und unbestimmter Artikel des Deutschen in ein entsprechendes Bestimmtheitsmerkmal am Regens. Eine größere Umgestaltung entsteht durch das Bestreben, die Präpositionen aus der syntaktischen Struktur zu entfernen. Hierbei wird eine Aufspaltung verschiedener Dependenztypen der Attribute und freien Angaben in ungefähr so viele Typen erforderlich, wie es Präpositionen gibt. Die Komplexität wird also nicht geringer; falls jedoch ein erheblicher Vorteil beim semantischen Parsing erzielt werden kann, mag man sich für eine solche Lösung entscheiden. Sie läge durchaus im Sinne Tesnières, der Präpositionen eine Translationsfunktion zuweist und sie daher nicht als selbstständige Knoten im Stemma repräsentiert (Tesnière 1959/1982, 366–369, vgl. Vennemann 1977, 273).

Die Frage der Behandlung von Funktionswörtern berührt die grundlegende Entscheidung, inwieweit die Syntax, die einen Entscheidungsraum für inhaltliche Mechanismen eröffnet, die semantischen Relationsmuster direkt widerspiegeln sollte. Manchen Autoren schwebt eine nur aus Inhaltselementen und ihren Relationen zusammengesetzte semantiknahe Struktur vor, so etwa Paul Garde (1977, 7). Die Entscheidung für eine Baumstruktur anstelle komplexerer Graphen (vgl. 3.1) bringt es allerdings mit sich, dass nicht jede semantische Prädikationsbeziehung im Satz auch einer direkten syntaktischen Relation entsprechen kann. Dies gilt insbesondere in Fällen wie denen, die bisweilen als doppelte Abhängigkeit (Mel'čuk 1988, 24–26; Wegener 1990, 157) oder mehrfache Regentien (Fraser 1989, 298: „multiple heads“) bezeichnet werden. Mel'čuk, der die doppelte Abhängigkeit diskutiert und für sein Modell verwirft, verwendet die Beispielsätze engl. *We found John working* oder russ. *My našli zal pustym* 'Wir fanden den Saal leer vor'. Hier besteht nach meiner Analyse eine direkte Regensrelation von dem finiten Verb zu jedem der drei Inhaltswörter (*we* Subjekt, *John* direktes Objekt, *working* Prädikativ; *my* Subjekt, *zal* direktes Objekt, *pustym* Prädikativ). Gleichzeitig enthalten die Sätze aber auch die Aussagen 'John arbeitet' und 'der Saal ist leer'. Das russische Beispiel zeigt zwischen *zal* und *pustym* überdies Genusreaktion – Tes-

nières (morphologisches) Kriterium für eine Dependenzrelation (vgl. 3.1).

Ich nehme Beispiele dieser Art jedoch nicht zum Anlass, die für die Formalisierung und die weitere Bearbeitung außerordentlich nützliche Beschränkung auf eine echte Baumstruktur aufzugeben, sondern überlasse es dem semantischen Modell, hier wie in einer Reihe anderer Fälle semantische Relationen zwischen syntaktisch nur mittelbar verbundenen Inhaltswörtern abzuleiten. Dieses Vorgehen entspricht der von Mel'čuk (1988, 49 Abb. 5) illustrierten Einsicht, nach der die Repräsentation von Sätzen auf der phonetischen und der morphologischen Ebene als lineare Kette, auf der syntaktischen als Baum und auf der semantischen als beliebig querverbundenes Netz erfolgen kann. Festzuhalten bleibt jedoch, dass die von dem hier beschriebenen dependenzsyntaktischen Modell geschaffenen Strukturen einem semantischen Modell sehr weit entgegenkommen, das von den inhaltlichen Aussagebeziehungen ausgeht, die sich direkt aus der syntaktischen „Konnexität“ ablesen lassen (Ajdkiewicz's Terminus, vgl. 3). Das oben erwähnte Modell Sadlers geht mit seiner Elementarstruktur X R Y (Regens – Relator – Dependens) diesen Weg. Klaus Heger macht den Rückgriff auf Prädikationsrelationen noch deutlicher, indem er in seiner Elementarstruktur A R P (Aktant – Relator – Prädikator) ausdrücklich von einem Prädikator spricht (Heger/Mudersbach 1984, 13). Heger bezieht seine so definierte Struktur direkt auf die von Tesnière in Stemmas dargestellte (Heger 1991, 45). In ganz anderer Form gehen zahlreiche japanische und chinesische Projekte direkt von einer Dependenzsyntax zur Wissensrepräsentation (z. B. Muraki/Ichijama/Fukumochi 1985; Yuan/Huang/Pan 1992). Charles Fillmore's Kasusgrammatik ist ursprünglich als semantisches Gegenstück zu einer Dependenzsyntax entworfen (Fillmore 1987) und Roger Schanks (1972) Modell der Conceptual Dependency baut auf inhaltlicher Ebene Strukturen auf, die den syntaktischen Dependenzlinien präzisierender Wörter in ähnlicher Weise folgen wie bei Sadler und Heger.

In den Anfangsjahren der Computerlinguistik wurde sehr lebhaft über die Frage diskutiert, ob dependenzielle oder konstituenzielle Grammatikmodelle es erlauben, *projektive* Strukturen zu erzeugen (vgl. Padučeva 1964; Marcus 1965). Kernpunkt dieser Diskussion ist die Frage, ob es möglich ist, eine Abbildung von Sätzen auf Baumstrukturen zu

definieren, die gleichzeitig die syntaktische Zusammengehörigkeit (Konstituenz) beziehungsweise die syntaktische Abhängigkeit (Dependenz) einerseits und die Wortfolge andererseits repräsentiert. Eine Baumstruktur kann jedoch nicht zwei Phänomene gleichzeitig wiedergeben, wenn es sich um voneinander unabhängige Phänomene handelt. Wenn gezeigt werden kann, dass in einer bestimmten Sprache Dependenz und Wortfolge jedes beliebigen Satzes projektiv dargestellt werden können, dann hängt die eine Erscheinung von der anderen ab. Wenn also Projektivität möglich ist, ist das eine der beiden Phänomene redundant und braucht nicht repräsentiert zu werden. Ist es dagegen nicht redundant, kann Projektivität nicht in allen Fällen gewährleistet werden. Die Baumstrukturen des Modells Metataxe repräsentieren daher nur die Dependenz. Wo die Wortfolge in der Anordnung der Dependenz von links nach rechts wiedergegeben werden kann, hilft dies der Intuition der Grammatiker und sollte daher nicht zu Gunsten einer normierten Dependenzabfolge aufgegeben werden. Wo jedoch Dependenz und Wortfolge einander entgegenstehen, spiegelt der Baum die lineare Wortfolge nicht wider.

Zwei weitere Fragen, in deren Beantwortung sich Grammatikmodelle auch innerhalb der dependenziellen Richtung unterscheiden, sind die nach abstrakten und leeren Knoten. Konstituenzgrammatiken bauen über dem Satz eine Struktur aus abstrakten Knoten auf, wodurch eine Unterscheidung zwischen terminalen und nichtterminalen Knoten entsteht. Die terminalen Knoten (oder Blätter) tragen Wörter, die nichtterminalen tragen abstrakte Kategorieangaben. Von der Anlage der Dependenzsyntax her wird eine solche Unterscheidung nicht gemacht. Da diese Modelle die Relationen zwischen Wörtern beschreiben, tragen im Prinzip alle Knoten Wörter und es gibt keine abstrakten Knoten. (Matthews' 1981, 81–82 Dependenzbäume, die an Robinsons 1969 anschließen, fallen hier aus dem Rahmen.) Hin und wieder werden jedoch Überlegungen diskutiert, die bestimmte Dependenzien und bestimmte Eigenschaften des Satzes nicht dessen innerem Regens (also meist dem finiten Verb), sondern einem abstrakten Knoten zuordnen. So setzt Eroms (1985, 309) einen abstrakten Knoten S an, der unter anderem den Aussagetyt des Satzes repräsentiert, wahrt jedoch das Prinzip, indem er auf diesen Knoten das Satzzeichen setzt und es zum „Wortäquivalent“ er-

klärt. Auch das Modell Metataxe verwendet Satzzeichen als Wörter, nimmt jedoch das Satzzeichen am Ende des Satzes als Dependenz des Hauptregens an. In Nebenordnungsstrukturen erscheint ein Komma in der Funktion einer Konjunktion (vgl. 3.7). Abstrakte Knoten kennt das Modell Metataxe daher nicht.

Leere Knoten werden bisweilen dort eingesetzt, wo der Grammatiker meint, dass ein Wort ausgelassen sei. Hier geht es nicht so sehr um Fälle wie Mel'čuks Satz russ. *Ivan učěnyj* 'Ivan ist Wissenschaftler' (Mel'čuk 1988, 15), in denen der Indikativ Präsens des Verbs *byl* 'sein' eine Leerstelle (in anderen Konstruktionen auch ein Gedankenstrich) ist. Da Mel'čuks Bäume die Wörter in der Grundform tragen, unterscheidet sich dieser Fall nicht einmal von anderen Verben. Von zentralerer Bedeutung sind elliptische Konstruktionen (vgl. 3.7).

## 6. Fazit

Das Grammatikmodell Metataxe umfasst die einsprachige Syntax der Analyse und Synthese sowie die zweisprachige Syntax der Strukturübertragung im maschinellen Übersetzungsvorgang. Sie ist modular konzipiert, insbesondere im Hinblick auf das Miteinander von Syntax und Wörterbuch, Strukturübertragung und lexikaler Übertragung sowie auch auf die Übertragbarkeit des Modells auf typologisch unterschiedliche Sprachen. Es steht überdies vor dem Hintergrund einer Auffassung vom modularen sprachtechnologischen Arbeiten auf den Ebenen der Linguistik, der Formalisierung und der Implementierung. Das Modell ist darauf ausgerichtet, mit syntaktischen Mitteln Entscheidungsräume zu eröffnen, in denen inhaltsgesteuerte Entscheidungsmechanismen arbeiten können. Es erlaubt einen unaufwendigen Übergang zu geeigneten semantischen und wissens-technischen Modellen.

*Ich danke Igor' Mel'čuk für konstruktive Kritik an einer Manuskriptfassung dieses Beitrags. Man rechne ihm nicht an, dass ich nicht alle seiner Hinweise übernommen habe.*

## 7. Literatur in Auswahl

- Ajdukiewicz, Kazimierz (1935): Die syntaktische Konnexität. In: *Studia Philosophica* 1, 1–27.  
 Badia, Toni (1993): Dependency and Machine Translation. In: Eynde, Frank Van (Hg.) (1993):

- Linguistic Issues in Machine Translation*. London/ New York, 105–137.
- Baumgärtner, Klaus (1970): Konstituenz und Dependenz. In: Steger, Hugo (Hg.) (1970): *Vorschläge für eine strukturelle Grammatik des Deutschen*. Darmstadt, 52–77.
- Bechraoui, Mohamed Fadhel (1992): La syntaxe de Tesnière: analyse métathéorique. In: *Linguistica Investigationes* 16 [1], 1–20.
- Dasgupta, Probal (1989): A Dependency Syntax of Bangla. In: Maxwell/Schubert (Hgg.) (1989), 89–113.
- Dolinina, Inga B. (1969): Sposoby predstavlenija sintaksičeskoj struktury predloženiya. In: Cholodovič, Aleksandr A. (Hg.) (1969): *Tipologija kauzativnyh konstrukcij*. Leningrad, 294–310.
- Engel, Ulrich (1977): *Syntax der deutschen Gegenwartssprache*. 3. Aufl. 1994. Berlin.
- Engel, Ulrich (1980): Fügungspotenz und Sprachvergleich. In: *Wirkendes Wort* 30, 1–22.
- Eroms, Hans-Werner (1985): Eine reine Dependenzgrammatik für das Deutsche. In: *Deutsche Sprache* 13, 306–326.
- Fillmore, Charles J. (1987): A Private History of the Concept „Frame“. In: Dirven, René/Radden, Günter (Hgg.) (1987): *Concepts of Case*. Tübingen, 28–36.
- Fraser, Norman (1989): Parsing and Dependency Grammar. In: *UCL Working Papers in Linguistics* 1, 296–319.
- Garde, Paul (1977): Ordre linéaire et dépendance syntaxique. In: *Bulletin de la Société de Linguistique de Paris* 72, 1–19.
- Gréciano, Gertrud/Schumacher, Helmut (Hgg.) (1996): *Lucien Tesnière – Syntaxe structurale et opérations mentales*. Tübingen.
- Groot, Albert W. de (1949): *Structurele syntaxis*. Den Haag.
- Hays, David G. (1963): Research Procedures in Machine Translation. In: Garvin, Paul L. (Hg.) (1963): *Natural Language and the Computer*. New York usw., 183–214.
- Hays, David G. (1964): Dependency Theory: A Formalism and Some Observations. In: *Language* 40, 511–525.
- Heger, Klaus (1991): Vom Stemma zum Aktantenmodell. In: Koch/Krefeld (Hgg.) (1991), 41–49.
- Heger, Klaus/Mudersbach, Klaus (1984): *Aktantenmodelle. Abhandlungen der Heidelberger Akademie der Wissenschaften, Philosophisch-historische Klasse. 1984/4*. Heidelberg/Birkenau.
- Hellwig, Peter (1986): Dependency Unification Grammar. In: *11th International Conference on Computational Linguistics*. Bonn, 195–198.
- Hellwig, Peter (1989): Parsing natürlicher Sprachen: Grundlagen. In: Bátor, István S./Lenders, Winfried/Putschke, Wolfgang (Hgg.) (1989): *Computational Linguistics / Computerlinguistik*. Berlin/ New York, 348–377.
- Hjelmslev, Louis (1943): *Omkring sprogteoriens grundlæggelse*. Neue Aufl. 1966. København.
- Hjelmslev, Louis (1963): *Sproget*. 3. Aufl. København.
- Hudson, Richard (1984): *Word Grammar*. Oxford/ New York.
- Hutchins, W. John (1988): Recent Developments in Machine Translation. In Maxwell/Schubert/Witkam (Hg.) (1988), 7–63.
- Iomdin, Leonid Lejbovič (1990): *Avtomatičeskaja obrabotka teksta na estestvennom jazyke: model' soglasovanija*. Moskva.
- Iordanskaja, Lidija Nikolaevna (1964): Svojtva pravil'noj sintaksičeskoj struktury i algoritmy ee obnaruženija. In: *Problemy kibernetiki* [11], 215–244.
- Joshi, Aravind K./Schabes, Yves (1991): *Tree-Adjoining Grammars and Lexicalized Grammars*. Philadelphia.
- Kacnel'son, Solomon Davidovič (1948): O grammatičeskoj kategorii. In: *Vestnik Leningradskogo Universiteta, Serija istorii, jazyka i literatury* [2], 114–134.
- Kacnel'son, Solomon Davidovič (1987): K ponjatiju tipov valentnosti. In: *Voprosy jazykoznanija* [3] – Übersetzung (1988): Zum Verständnis von Typen der Valenz. In: *Sprachwissenschaft* 13, 1–30.
- Koch, Peter (1996): La métatase actantielle. In: Gréciano/Schumacher (Hg.) (1996), 211–224.
- Koch, Peter/Krefeld, Thomas (1991): Dependenz und Valenz in romanischen Sprachen. In: Koch/Krefeld (Hgg.) (1991), 5–38.
- Koch, Peter/Krefeld, Thomas (Hgg.) (1991): *Connexiones Romanicae*. Tübingen.
- Landsbergen, Jan (1987): Isomorphic Grammars and Their Use in the Rosetta Translation System. In: King, Margaret (Hg.) (1987): *Machine Translation Today*. Edinburgh, 351–372.
- Landsbergen, Jan (1994): A Compositional Definition of the Translation Relation. In: Appelo, Lisette u. a. (Hgg.) (1992): *Compositional Translation by M.T. Rosetta*. Dordrecht/Boston/London, 15–33, 439–455.
- Lappin, Shalom/McCord, Michael (1990): Anaphora Resolution in Slot Grammar. In: *Computational Linguistics* 16, 197–212.
- Lobin, Henning (1989): A Dependency Syntax of German. In: Maxwell/Schubert (Hgg.) (1989), 17–38.
- Marcus, Solomon (1965): Sur la notion de projectivité. In: *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 11, 181–192.
- Matthews, Peter H. (1981): *Syntax*. Cambridge usw.

- Maxwell, Dan (1989): English-Esperanto Metataxis. In: Maxwell/Schubert (Hgg.) (1989), 267–297.
- Maxwell, Dan/Schubert, Klaus (Hgg.) (1989): *Metataxis in Practice*. Dordrecht/Providence.
- Maxwell, Dan/Schubert, Klaus/Witkam, Toon (Hgg.) (1988): *New Directions in Machine Translation*. Dordrecht/Providence.
- Mel'čuk, Igor' A. (1964): *Avtomatičeskij sintaksičeskij analiz*. Bd. 1. Novosibirsk.
- Mel'čuk, Igor' A. (1988): *Dependency Syntax: Theory and Practice*. Albany.
- Muraki, Kazunori/Ichijama, Shunji/Fukumochi, Yasutomo (1985): Augmented Dependency Grammar. In: *Second Conference of the European Chapter of the Association for Computational Linguistics*, 198–204.
- Nichols, Johanna (1986): Head-Marking and Dependent-Marking Grammar. In: *Language* 62, 56–119.
- Padučeva, Elena Viktorovna (1964): O sposobach predstavljenija sintaksičeskoj struktury predložennija. In: *Voprosy jazykoznanija* 13 [2], 99–113.
- Papegaaï, Bart/Schubert, Klaus (1988): *Text Coherence in Translation*. Dordrecht/Providence.
- Prószéky, Gábor/Koutny, Ilona/Wacha, Balázs (1989): A Dependency Syntax of Hungarian. In: Maxwell/Schubert (Hgg.) (1989), 151–181.
- Rasmussen, Michael (1992): *Hjemslevs sprogteori*. Odense.
- Robinson, Jane J. (1969): A Dependency-Based Transformational Grammar. In: *Proceedings of the 10th International Congress of Linguists*. Bd. 2. București, 807–813.
- Sadler, Victor (1989): *Working with Analogical Semantics*. Dordrecht/Providence.
- Sadler, Victor (1991): Machine Translation Project Reaches Watershed. In: *Language Problems and Language Planning* 15, 78–83.
- Sato, Shigeru (1989): A Dependency Syntax of Japanese. In: Maxwell/Schubert (Hgg.) (1989); 183–206.
- Schank, Roger C. (1972): Conceptual Dependency. In: *Cognitive Psychology* 3, 552–631.
- Schmidt, Jürgen Erich (1991): Konvergenzen zwischen neueren Grammatiktheorien und Deskriptionsgrammatiken? In: Feldbusch, Elisabeth/Pogarell, Reiner/Weiß, Cornelia (Hgg.) (1991): *Neue Fragen der Linguistik*. Bd. 1. Tübingen, 211–218.
- Schubert, Ingrid (1989): A Dependency Syntax of Danish. In: Maxwell/Schubert (Hgg.) (1989), 39–67.
- Schubert, Klaus (1986): *Syntactic Tree Structures in DLT*. Utrecht.
- Schubert, Klaus (1987): *Metataxis*. Dordrecht/Providence.
- Schubert, Klaus (1988a): Modularity in Linguistic System Design. In: *Zbornik IV. konference Računalniška obdelava jezikovnih podatkov*. [Ljubljana], 91–106.
- Schubert, Klaus (1988b): The Architecture of DLT – Interlingual or Double-Direct? In: Maxwell/Schubert/Witkam (Hgg.) (1988), 131–144.
- Schubert, Klaus (1988c): Implicitness as a Guiding Principle in Machine Translation. In: Vargha, Dénes (Hg.) (1988): *Coling Budapest*. Budapest, 599–601.
- Schubert, Klaus (1988d): Ausdruckskraft und Regelmäßigkeit. In: *Language Problems and Language Planning* 12, 130–147 – Aktualisierte Fassung in: Tonkin, Humphrey (Hg.) (1997): *Esperanto, Interlinguistics, and Planned Language*. Lanham/New York/Oxford, 117–139.
- Schubert, Klaus (1989a): Dependency Syntax for Parsing and Generation. In: Maxwell/Schubert (Hgg.) (1989), 7–15.
- Schubert, Klaus (1989b): A Dependency Syntax of Esperanto. In: Maxwell/Schubert (Hgg.) (1989), 207–232.
- Schubert, Klaus (1989c): The Theory and Practice of Metataxis. In: Maxwell/Schubert (Hgg.) (1989), 233–245.
- Schubert, Klaus (1992a): Esperanto as an Intermediate Language for Machine Translation. In: Newton, John (Hg.) (1992): *Computers in Translation*. London/New York, 78–95.
- Schubert, Klaus (1992b): Diederichsen, Tesnière och datorn. In: Louis-Jensen, Jonna/Poulsen, Johan Hendrik W. (Hgg.) (1992): *The Nordic Languages and Modern Linguistics* 7. Tórshavn, Bd. 2, 565–577.
- Schubert, Klaus (1993): Semantic Compositionality. In: *Linguistics* 31, 311–365.
- Schubert, Klaus (1996): Lexikon und grammatischer Regelmechanismus in der maschinellen Übersetzung. In: Walczak, R. (Hg.) (1996): *Forschungsbericht 1995/96*. Flensburg, 97–107.
- Schubert, Klaus (1997): Remote-Access Translation Services. In: *Translating and the Computer* 19. London [separat paginiert].
- Schubert, Klaus (demn.): Maschinelle Übersetzung mit Dependenzgrammatiken [Arbeitstitel]. Art. Nr. 84 in Band 2 dieses Handbuchs.
- Stone, Matthew/Doran, Christine (1997): Sentence Planning as Descripton Using Tree Adjoining Grammar. In: *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, 198–205.
- Świdziński, Marek (1989): A Dependency Syntax of Polish. In: Maxwell/Schubert (Hgg.) (1989), 69–87.
- Tamis, Dorine (1989): Esperanto-French Metataxis. In: Maxwell/Schubert (Hgg.) (1989), 247–266.
- Tarvainen, Kalevi (1981): *Einführung in die Dependenzgrammatik*. Tübingen.

Tarvainen, Kalevi (1985): *Kielioppia kontrastiivisesti*. Jyväskylä.

Tarvainen, Kalevi (1989): A Dependency Syntax of Finnish. In: Maxwell/Schubert (Hgg.) (1989), 115–150.

Tesnière, Lucien (1959): *Éléments de syntaxe structurale*. 2. Aufl., 4. Druck 1982. Paris.

Vennemann, Theo (1977): Konstituenz und Dependenz in einigen neueren Grammatiktheorien. In: *Sprachwissenschaft* 2, 259–301.

Wegener, Heide (1990): Komplemente in der Dependenzgrammatik und in der Rektions- und Bindungstheorie. In: *Zeitschrift für germanistische Linguistik* 18, 150–184.

Werner, Edeltraud (1993): *Translationstheorie und Dependenzmodell*. Tübingen/Basel.

Witkam, Toon (1983): *Distributed Language Translation*. Utrecht.

Yuan, Chunfa/Huang, Changning/Pan, Shimei (1992): Knowledge Acquisition and Chinese Parsing Based on Corpus. In: Boitet, Christian (Hg.): *Proceedings of the Fifteenth International Conference on Computational Linguistics*, 1300–1304.

Zuijlen, Job M. van (1989): Aspects of Metataxis. Formalization. In: Maxwell/Schubert (Hgg.) (1989), 299–320.

Zuijlen, Job M. van (1992): Complexiteit en automatisch vertalen. In: Alkemade, M. J. A. (Hg.) (1992): *Inspelen op complexiteit*. Alphen aan den Rijn, 220–228.

Klaus Schubert, Flensburg (Deutschland)

## 46. Dependency, Valency and Head-Driven Phrase-Structure Grammar

1. Introduction and Remarks on Formal Background
2. General Linguistic Background
3. Head-Drivenness and Dependency
4. Subcategorization or Valency
5. Interplay of Dependency, Valency and Word Order
6. Long Distance Dependency and Non-Projectivity
7. Conclusions
8. Select Bibliography

### 1. Introduction and Remarks on Formal Background

Head-driven Phrase Structure Grammar (HPSG, Pollard and Sag, 1987, 1994) is a nontransformational phrase-structure-based theory representing syntactic categories by *feature structures* (*feature bundles*). A feature-structure is to be understood as a *description of properties of an object* – that is, any object with such properties is considered to be described by the feature structure. In this sense, a feature structure *defines the set of all objects it describes*. Note that the descriptions are purely static – in particular, no changes performed on objects (transformations) are employed. This means that HPSG can be grouped together with other purely declarative approaches, that is, among approaches for which in theory (as frequently opposed to practice) the contingent language processing, such as parsing or generation, is independent from the language description.

The basic operation with feature structures is *unification* (cf. Chapter 44 Dependency Unification Grammar). Further, in HPSG an extremely important role is played by *functional constraints*, that is, by relations binding together the values of different features of a feature structure. As a trivial, non-linguistic example of a functional constraint, one might consider the relations between the value of the feature *surface* and values of other features within the feature-structure representations of a (general) quadrangle and a (general) circle in the following feature structures

$$\left[ \begin{array}{l} \textit{type\_of\_object} : \textit{quadrangle} \\ \textit{shorter\_side} : X \\ \textit{longer\_side} : Y \\ \textit{surface} : X*Y \end{array} \right]$$

$$\left[ \begin{array}{l} \textit{type\_of\_object} : \textit{circle} \\ \textit{radius} : R \\ \textit{surface} : \pi*R*R \end{array} \right]$$

A certain particularity of HPSG consists in the fact that truly all linguistic knowledge is to be expressed in the feature structure format: we shall discuss some more examples below, but even here it is necessary to say that even the syntactic structures employed (that is, traditional constituent trees) are formally coded as feature structures, in particular, using the feature dtrs (and others), e. g., the tree from (1) would be coded in HPSG-like notation as in (2), where daughters are given as elements of a list.

Dependenz und Valenz  
Dependency and Valency

HSK 25.1



# Handbücher zur Sprach- und Kommunikations- wissenschaft

Handbooks of Linguistics  
and Communication Science

Manuels de linguistique et  
des sciences de communication

Mitbegründet von Gerold Ungeheuer (†)  
Mitherausgegeben 1985–2001 von Hugo Steger

Herausgegeben von / Edited by / Edités par  
Herbert Ernst Wiegand

Band 25.1

Walter de Gruyter · Berlin · New York  
2003

# Dependenz und Valenz Dependency and Valency

Ein internationales Handbuch der  
zeitgenössischen Forschung  
An International Handbook of  
Contemporary Research

Herausgegeben von / edited by  
Vilmos Ágel, Ludwig M. Eichinger, Hans-Werner  
Eroms, Peter Hellwig, Hans Jürgen Heringer,  
Henning Lobin

1. Halbband / Volume 1

Walter de Gruyter · Berlin · New York  
2003

⊗ Printed on acid-free paper which falls within the guidelines  
of the ANSI to ensure permanence and durability.

ISBN 3-11-014190-6

*Bibliografische Information der Deutschen Bibliothek*

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen  
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet  
über <<http://dnb.ddb.de>> abrufbar.

© Copyright 2003 by Walter de Gruyter GmbH & Co. KG, 10785 Berlin, Germany.  
All rights reserved, including those of translation into foreign languages. No part of this book may be  
reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy,  
recording or any information storage and retrieval system, without permission in writing from the publisher.  
Printed in Germany  
Typesetting: META-Systems GmbH, Wustermark  
Binding: Lüderitz & Bauer-GmbH, Berlin  
Coverdesign: Rudolf Hübler, Berlin

## Inhalt / Contents

### 1. Halbband / Volume 1

Vorwort . . . . .	XI
Preface . . . . .	XV

#### I. Das Dependenz- und Valenzparadigma in Natur- und Geisteswissenschaften The Dependency and Valency Paradigm in the Arts and Sciences

1. <i>Klaus Mainzer</i> , Das Abhängigkeitskonzept in den Wissenschaften . . . . .	1
2. <i>Stefan Pongó</i> , Die Wertigkeitsmetapher . . . . .	7
3. <i>Kurt Otto Seidel</i> , Valenzverwandte Ansätze in der Antike . . . . .	14
4. <i>Lauri Seppänen</i> , Mit der Valenz verwandte Begriffe im Mittelalter: ein Überblick . . . . .	20
5. <i>Jonathan Owens</i> , Valency-like Concepts in the Arabic Grammatical Tradition . . . . .	26
6. <i>Dieter Cherubim</i> , Valenzverwandte Ansätze in Humanismus und Aufklärung: ein Überblick . . . . .	32
7. <i>Klaus Mudersbach</i> , Mathematische und logische Rekonstruktion des Abhängigkeits- und Valenz-Konzepts . . . . .	37

#### II. Lucien Tesnière und seine Zeit Lucien Tesnière and his Times

8. <i>Jean Fourquet</i> , Lucien Tesnière. Ein Zeugnis 1933–1993 . . . . .	67
9. <i>Hans Jürgen Heringer</i> , Lucien Tesnière. Sein Leben . . . . .	70
10. <i>John Ole Askedal</i> , Das Valenz- und Dependenzkonzept bei Lucien Tesnière . . . . .	80
11. <i>Bernhard Engelen</i> , Die Wortartenlehre bei Lucien Tesnière . . . . .	100
12. <i>Gerd Wotjak</i> , Zu Tesnières Semantikkonzept . . . . .	108
13. <i>Edeltraud Werner</i> , Das Translationskonzept Lucien Tesnières . . . . .	115
14. <i>Hans Jürgen Heringer</i> , Die Junktionstheorie Lucien Tesnières . . . . .	129
15. <i>Richard Waltereit</i> , Negation und Frage bei Lucien Tesnière . . . . .	139
16. <i>Peter Koch</i> , Metataxe bei Lucien Tesnière . . . . .	144
17. <i>Hans-Werner Eroms</i> , Die Wegbereiter einer deutschen Valenzgrammatik . . . . .	159

<b>III. Dependenz. Grundlagen und Grundfragen</b>	
<b>Dependency. Basic Principles and Basic Issues</b>	
18.	<i>Willy Van Langendonck</i> , The Dependency Concept and its Foundations . . . . . 170
19.	<i>Igor Mel'čuk</i> , Levels of Dependency Description: Concepts and Problems . . . . . 188
20.	<i>Pál Uzonyi</i> , Dependenzstruktur und Konstituenzstruktur . . . . . 230
21.	<i>Hans-Werner Eroms</i> , <i>Hans Jürgen Heringer</i> , Dependenz und lineare Ordnung . . . . . 247
22.	<i>Peter Colliander</i> , Dependenzstruktur und grammatische Funktion . . . 263
23.	<i>Stanley Starosta</i> , Dependency Grammar and Lexicalism . . . . . 270
24.	<i>Wha-Young Jung</i> , Rektion und Kongruenz in der Dependenzgrammatik . . . . . 282
25.	<i>Norbert Bröker</i> , Formal Foundations of Dependency Grammar . . . . . 294
26.	<i>Elisabeth Leiss</i> , Empirische Argumente für Dependenz. . . . . 311
27.	<i>Henning Lobin</i> , Dependenzgrammatik und Kategorialgrammatik . . . . 325
28.	<i>Thomas Michael Groß</i> , Dependency Grammar's Limits – and Ways of Extending Them . . . . . 331
<b>IV. Valenz. Grundlagen und Grundfragen</b>	
<b>Valency. Basic Principles and Basic Issues</b>	
29.	<i>Gisela Zifonun</i> , Grundlagen der Valenz . . . . . 352
30.	<i>Joachim Jacobs</i> , Die Problematik der Valenzebenen . . . . . 378
31.	<i>Hans-Joachim Meinhard</i> , Ebenen der Valenzbeschreibung: Die logische und die semantische Ebene . . . . . 399
32.	<i>Norbert Richard Wolf</i> , Ebenen der Valenzbeschreibung: Die syntaktische Ebene . . . . . 404
33.	<i>Rüdiger Harnisch</i> , Ebenen der Valenzbeschreibung: die morphologische Ebene . . . . . 411
34.	<i>Christina Gansel</i> , Valenz und Kognition . . . . . 422
35.	<i>Viktor S. Xrakovskij</i> , Valenz und Sprachtypologie. . . . . 444
36.	<i>Charles Fillmore</i> , Valency and Semantic Roles: the Concept of Deep Structure Case . . . . . 457
37.	<i>Klaus Welke</i> , Valenz und semantische Rollen: das Konzept der Theta-Rollen . . . . . 475
38.	<i>Marcel Vuillaume</i> , Valenz und Satzbauplan . . . . . 484
39.	<i>Henrik Nikula</i> , Valenz und Pragmatik . . . . . 499
<b>V. Dependenzuelle Theorien</b>	
<b>Dependency Theories</b>	
40.	<i>Richard Hudson</i> , Word Grammar . . . . . 508
41.	<i>Stanley Starosta</i> , Lexicase Grammar . . . . . 526
42.	<i>Sylvain Kahane</i> , The Meaning-Text Theory . . . . . 546
43.	<i>Eva Hajičová</i> , <i>Petr Sgall</i> , Dependency Syntax in Functional Generative Description . . . . . 570

44.	<i>Peter Hellwig</i> , Dependency Unification Grammar . . . . .	593
45.	<i>Klaus Schubert</i> , Metataxe: ein Dependenzmodell für die computerlinguistische Praxis . . . . .	636
46.	<i>Karel Oliva</i> , Dependency, Valency and Head-Driven Phrase-Structure Grammar . . . . .	660
47.	<i>Srinivas Bangalore, Aravind K. Joshi, Owen Rambow</i> , Dependency and Valency in Other Theories: Tree Adjoining Grammar . . . . .	669
48.	<i>Dan Maxwell</i> , The Concept of Dependency in Morphology . . . . .	678
49.	<i>Henning Lobin</i> , Konzeptuelle Semantik als dependenzielle Theorie . . .	684
50.	<i>Wolfgang Menzel</i> , Semantische Netze und Dependenzgrammatik . . . .	691
51.	<i>Claudia Villiger</i> , Dependenzielle Textmodelle . . . . .	703

VI. Valenz: Schwerpunkte der Forschung  
Valency: Core Research Areas

52.	<i>Marja Järventausta</i> , Das Verb als strukturelles Zentrum des Satzes . . .	717
53.	<i>Irma Hyvärinen</i> , Der verbale Valenzträger . . . . .	738
54.	<i>Angelika Storrer</i> , Ergänzungen und Angaben . . . . .	764
55.	<i>Marja Järventausta</i> , Das Subjektproblem in der Valenzforschung . . . .	781
56.	<i>Heinz Vater</i> , Valency Potency and Valency Realization . . . . .	794
57.	<i>Fritz Pasierbsky</i> , Toward a Classification of Complements . . . . .	803
58.	<i>Kjell Johan Sæbo</i> , Valency and Context Dependence . . . . .	814
59.	<i>Wolfgang Teubert</i> , Die Valenz nichtverbaler Wortarten: das Substantiv	820
60.	<i>Thomas Michael Groß</i> , The Valency of Non-Verbal Word Classes: the Adjective . . . . .	835
61.	<i>Gertrud Gréciano</i> , Probleme der Valenz in der Phraseologie . . . . .	843

2. Halbband (Überblick über den vorgesehenen Inhalt) /  
Volume 2 (Preview of Contents)

VII. Grammatische Phänomene unter Dependenz- und  
Valenzgesichtspunkten  
Grammatical Phenomena as Seen from Dependency and  
Valency Viewpoints

62.	<i>Ludwig M. Eichinger</i> , Wortstellung: die verbalen Teile – Interne Valenzen und Klammerhierarchie
63.	<i>Ursula Hoberg</i> , Wortstellung: valenzgebundene Teile und Positionspräferenzen
64.	<i>Erwin Koller</i> , Wortstellung: textfunktionale Kriterien
65.	<i>John Ole Askedal</i> , Infinitivkonstruktionen
66.	<i>Wilhelm Oppenrieder</i> , Subjekt- und Objektsätze
67.	<i>Henrik Nikula</i> , Unpersönliche Konstruktionen
68.	<i>Renate Pasch</i> , Adverbial- und Relativsätze
69.	<i>Eva Breindl-Hiller</i> , Präpositionalphrasen
70.	<i>Josef Schmid</i> , Die „freien“ Dative
71.	<i>Roman Sadziński</i> , Konversen und Diathesen

- 72. *Henning Lobin*, Koordination in Dependenzgrammatiken
- 73. *Wilfried Kürschner*, Negation in Dependenzgrammatiken
- 74. *Thomas A. Fritz*, Modalität in Dependenzansätzen
- 75. *Hans-Werner Eroms*, Diskurspartikeln und Modalwörter
- 76. *Jürgen Erich Schmidt*, Serialisierung in der Nominalphrase
- 77. *Stefan Schierholz*, Flexion in der Nominalphrase
- 78. *Ludwig M. Eichinger*, Dependenz in der Wortbildung

### VIII. Dependenz in der maschinellen Sprachverarbeitung Dependency in Computer-Based Language Processing

- 79. *Peter Hellwig*, Parsing with Dependency Grammars
- 80. *Helmut Horacek*, Generierung mit Dependenzgrammatiken
- 81. *Klaus Schubert*, Maschinelle Übersetzungen mit Dependenzgrammatiken

### IX. Dependenz und Valenz in der kontrastiven Linguistik Dependency and Valency in Contrastive Linguistics

- 82. *Marja-Leena Piitulainen*, Dependenz und Valenz in der kontrastiven Syntax: ein Überblick
- 83. *Rudolf Emons*, Kontrastive Fallstudie: Deutsch – Englisch
- 84. *Albrecht Plewnia*, Kontrastive Fallstudie: Deutsch – Französisch
- 85. *Teresa Bianco*, Kontrastive Fallstudie: Deutsch – Italienisch
- 86. *Christian Fandrych*, Kontrastive Fallstudie: Deutsch – Spanisch
- 87. *Norbert Nübler*, Kontrastive Fallstudie: Deutsch – Russisch
- 88. *Christoph Schatte*, Kontrastive Fallstudie: Deutsch – Polnisch
- 89. *Ulrich Engel*, Kontrastive Fallstudie: Deutsch – Serbisch/Kroatisch
- 90. *Emilia Baschewa-Monova*, Kontrastive Fallstudie: Deutsch – Bulgarisch
- 91. *Speranța Stănescu*, Kontrastive Fallstudie: Deutsch – Rumänisch
- 92. *Siamak Mohadjer-Ghomi*, Kontrastive Fallstudie: Deutsch – Persisch
- 93. *Irma Hyvärinen*, Kontrastive Fallstudie: Deutsch – Finnisch
- 94. *Jan Daugaard*, Contrastive Case Study: German – Danish
- 95. *Peter Bassola*, Kontrastive Fallstudie: Deutsch – Ungarisch
- 96. *Abdesrazzaq Msellek*, Kontrastive Fallstudie: Deutsch – Arabisch
- 97. *Han Wanbeng*, Kontrastive Fallstudie: Deutsch – Chinesisch
- 98. *Susumu Zaima*, Kontrastive Fallstudie: Deutsch – Japanisch
- 99. *Lie Kwang-Sook, Mi-Kyung Hong*, Kontrastive Fallstudie: Deutsch – Koreanisch

### X. Das Valenzkonzept in der Grammatikographie The Valency Concept in Grammaticographical Studies

- 100. *Ulrich Engel*, Das Valenzkonzept in der Grammatikographie: ein Überblick
- 101. *Dorothea Kobler-Trill, Anita Schilcher*, Das Valenzkonzept in Referenzgrammatiken: deutsche Schulgrammatiken und Sprachbücher

102. *Karl-Ernst Sommerfeldt*, Das Valenzkonzept in Referenzgrammatiken: Gebrauchsgrammatiken
103. *Elke Hentschel*, Das Valenzkonzept in Referenzgrammatiken: Handbuchgrammatiken
104. *Maria Thurmair*, Das Valenzkonzept in Referenzgrammatiken: Grammatiken für Deutsch als Fremdsprache
- XI. Das Valenzkonzept in der Lexikographie  
The Valency Concept in Lexicography
105. *Jaqueline Kubczak*, Valenzinformationen in allgemeinen einsprachigen Wörterbüchern
106. *Peter Bassola*, Valenzinformationen in allgemeinen zweisprachigen Wörterbüchern
107. *Helmut Schumacher*, Deutschsprachige Valenzwörterbücher
108. *Winfried Busse*, Valenzwörterbücher in anderen Sprachen
109. *Ludwig M. Eichinger*, Kontrastive zweisprachige Valenzwörterbücher
- XII. Das Valenzkonzept in der Sprachgeschichtsforschung:  
ausgewählte Bereiche  
The Valency Concept in Research into the History of  
Language: Selected Areas
110. *Hans Jürgen Heringer*, Prinzipien des Valenzwandels
111. *Jarmo Antero Korhonen*, Valenzwandel am Beispiel des Deutschen
112. *Albrecht Greule*, Historische Fallstudie: Althochdeutsch
113. *Hans-Joachim Solms*, Historische Fallstudie: Mittelhochdeutsch
114. *Jarmo Antero Korhonen*, Historische Fallstudie: Frühneuhochdeutsch
115. *Rosemarie Lühr*, Historische Fallstudie: Altsächsisch
116. *Herbert Schendl*, Historische Fallstudie: Alt- und Mittelenglisch
117. *Peter Stein, Claudia Benneckenstein*, Historische Fallstudie: Altfranzösisch
118. *Tamás Forgács*, Historische Fallstudie: Altungarisch
- XIII. Das Valenzkonzept in weiteren Forschungsbereichen  
The Valency Concept in Other Areas of Research
119. *Franz Simmler*, Varietätenlinguistik: Fachsprachen
120. *Bernhard Sowinski*, Varietätenlinguistik: Sprache der Literatur
121. *Franz Patocka*, Varietätenlinguistik: Dialekte
122. *Heidrun Gerzymisch-Arbogast*, Valenz und Übersetzung