

3. Darstellung des Programms nach Projektbereichen und Teilprojekten

3.0. Projektbereich A: Methodenorientierte Untersuchungen

A1: **Hinrichs/Kübler**: Repräsentation und Erschließung linguistischer Daten

A2: **Mönnich**: Linguistische Beschreibungsstrukturen als Datentypen

A3: **Sternefeld**: Suboptimale syntaktische Strukturen

A4: **Pafel**: On- und Offline-Aspekte semantischer Verarbeitung

A5: **Richter**: Distributionsidiosynkrasien in der logischen Form

Die im Sonderforschungsbereich angestrebte systematische Reflexion auf die Rolle verschiedener Datenklassen für die empirische Bestätigung linguistischer Theorien findet ihre komplementäre Ergänzung durch die Reflexion auf die theoriegeleitete Interpretation empirischer Daten. In dieser doppelten Perspektive spiegelt sich der Gebrauch des Begriffs „Datenstruktur“, mit dem einerseits generell auf den Gegenstandsbereich einer Theorie Bezug genommen werden kann, der andererseits in seiner technischen Verwendung auf eine abstrakte Charakterisierung der Struktur der Objekte verweist, die von der jeweiligen Theorie betrachtet werden. Während in der datenorientierten Perspektive der Hauptakzent auf den Vergleich und die Gewichtung von Datentypen in ihrer Relevanz für bestimmte Fragestellungen gelegt wird, zeichnet sich die theoriegeleitete Perspektive dadurch aus, dass sie ihre Objekte im Hinblick auf jene Eigenschaften idealisiert, die im Bereich des „Strukturhorizonts“ eines theoretischen Paradigmas liegen.

Die systematische Reflexion auf die Rolle verschiedener Datenklassen ist geeignet, die Frage nach dem linguistischen Erkenntnisobjekt in empirischer und theoretischer Perspektive gleichermaßen zu akzentuieren. Unter dieser doppelten Perspektive lassen sich die methoden- und korpusorientierten Untersuchungen linguistischer Datenstrukturen und Datenklassen des Projektbereichs A stichwortartig an den Problemen der Repräsentations- und Kodierungsverfahren, der Semantik abstrakter Datentypen und der scharfen theoriebasierten Klassifikation von Datenfamilien, die im Widerspruch zu unscharfen Bewertungsurteilen stehen, festmachen. In dieser Triade kommt die traditionelle Unterscheidung in formale, deskriptive und interpretatorische Aspekte einer linguistischen Metatheorie zum Ausdruck.

Das Verhältnis von Empirie und Theorie in der Grammatikforschung als zentraler Forschungsgegenstand des Sonderforschungsbereichs wird im Teilprojekt A1 am Datentyp von elektronisch verfügbaren Textkorpora des Deutschen und deren automatischer

Annotation untersucht. Die automatische linguistische Annotation dieses Datentyp erfordert in methodischer Hinsicht die Erweiterung traditioneller Grammatikformalismen um eine probabilistische Komponente bzw. den Einsatz maschineller Lernverfahren. Ein wesentliches Erkenntnisinteresse des Projekts A1 besteht darin, wie die Stärken derartiger Verfahren mit klassischen regelbasierten Ansätzen in hybriden Architekturen verknüpft werden können. In empirischer Hinsicht konzentriert sich A1 auf die Annotation komplexer Satzgefüge, und hier speziell auf die Phänomenbereiche: Erweiterung des Inventars grammatischer Funktionen um Adjunkte, Annotation von Parataxe, Hypotaxe und Koordinationsstrukturen, sowie Auflösung anaphorischer Beziehungen auf Satz- und Diskursebene. Die Ergebnisse der Annotation werden den Teilprojekten des SFB, bei denen die datenorientierte Untersuchung komplexer Satzgefüge im Vordergrund steht, zur Verfügung gestellt. In Kooperation mit dem Teilprojekt C1 stellt das Projekt A1 den anderen Einzelprojekten ausserdem Expertise bei der Enkodierung von Textkorpora in Standardformaten, speziell in den Auszeichnungssprachen XML und XHTML, zur Verfügung.

Linguistische Datenstrukturen lassen sich unter verschiedenen Gesichtspunkten zu Einheiten zusammenfassen. Daten, die zu einem Phänomenbereich gehören, die einem Areal zuzuordnen sind, die einer historischen Phase entstammen oder den Sprachgebrauch einer sozialen Schicht dokumentieren, sind z.B. als empirische Basis für die Verifikation oder die Falsifikation einer bestimmten Hypothese geeignet. Das Teilprojekt A2 abstrahiert von diesen inhaltlichen Gesichtspunkten und betrachtet stattdessen die strukturellen Gemeinsamkeiten, die für die implizite Gegenstandskonstitution ganzer Theorieparadigmen kennzeichnend sind. Dabei sollen in der beantragten Förderphase drei Schwerpunkte gebildet werden. In Fortführung einer längerfristig angelegten Untersuchung der adäquaten Modelltheorie für linguistische Theorien sollen modelltheoretische und komplexitätstheoretische Eigenschaften der GB-Theorie und der Optimalitätstheorie analysiert werden. Den Zweiten Schwerpunkt bildet eine vergleichende Evaluation der generativen Kapazität mild-kontextsensitiver Systeme. Schließlich sollen im dritten Schwerpunkt effiziente Abfragesprachen und Abfragewerkzeuge entwickelt werden, die einer Reihe von Projekten aus dem A- und B-Bereich zur Verfügung gestellt werden sollen.

Das Teilprojekt A3 untersucht marginalgrammatische und ungrammatische syntaktische Strukturen als Evidenz für syntaktische Theoriebildung. Suboptimale Beispiele werden nicht als periphere Ausnahmefälle, sondern als vollwertige Daten behandelt. Die bisherige Arbeit hat verdeutlicht, dass die Anerkennung der Realität von Gradienz sowohl für spezifische grammatische Phänomene wie auch für die breitere Theoriebildung von großem Interesse ist. Drei Arbeitsbereiche lassen sich unterscheiden. Im Bereich 1 werden aus der Literatur suboptimale Beispielstrukturen extrahiert und in einer Datenbank gespeichert; diese dient als Referenzquelle für einschlägige, in der Literatur erschienene Urteile, wobei auch unabhängig erhobene Daten als objektive Ver-

gleichsgrundlage dienen. Die Datenbank ist über Schlüsselwörter sowie über Strukturabfragen on-line recherchierbar. Die syntaktische Annotation folgt weitgehend einer theoretisch fundierten Merkmalsgrammatik. Im Bereich 2 werden unter streng kontrollierten Bedingungen introspektive Urteile zu ausgewählten Syntax-Themen experimentell erhoben und überprüft. Die primäre Zielsetzung ist syntax-theoretisch, ihre Durchführung ist psycholinguistisch-experimentell. Die bisherigen Untersuchungen von grammatischen Beschränkung und deren Interaktion führten zur Entwicklung des Decathlon-Modells, welches in der nächsten Periode geprüft und ergänzt werden muss, um als Grundlage für den Entwurf eines empirisch adäquaten Grammatikmodells dienen zu können. Arbeitsbereich 3 ist methodologischer Natur und vereint die beiden ersten. Hier geht es darum, eine Erhebungsmethode zu entwickeln, die als Standard für den Umgang mit gradierten Daten dienen soll. Aufbauend auf den Erkenntnissen zu Eigenschaften von relativen (d.h. nicht kategorischen) Urteilen erfolgt die Standardisierung komparativ zu Eichstrukturen, deren Beurteilungen experimentell festgestellt worden sind. Die kategorialen Urteile der A3-Datenbank werden unter Anwendung dieser Methodologie in relative, standardisierte gradierte Urteile überführt.

Die Datenerhebung in der semantischen Forschung ist durch eine starke Polarisierung gekennzeichnet: Auf der einen Seite werden semantische Theorien hauptsächlich auf der Grundlage semantischer Intuitionen gebildet, auf der anderen Seite gibt es in jüngster Zeit vermehrt Versuche, semantische Verarbeitung mit psycholinguistischen Methoden in den Griff zu bekommen. Unter diesem Spannungsbogen steht das Projekt A4, das sowohl methodologische als auch inhaltlich semantische Ziele verfolgt. Die erste Projektphase war hauptsächlich der methodologischen Fragestellung gewidmet, wie semantische Urteile (Off-Line) reliabel zu erfassen sind. Die von A4 evaluierten Methoden sollen auf neue inhaltliche Fragestellungen sowie auf eine weitere Modalität (gesprochene Sprache) angewandt werden. Dabei soll das Augenmerk in der kommenden Phase darauf gelegt werden, semantische Theorie und Aspekte semantischer Verarbeitung zusammenzubringen. Hierzu soll vermehrt von On-Line-Methoden Gebrauch gemacht werden, um so einen inhaltlichen Beitrag zu den Phänomenen I-Topikalisierung, Alternativensemantik sowie gebundenen Variablen zu leisten.

Das im Sonderforschungsbereich erreichte größere Maß an Empiriebezug der Grammatiktheorie einerseits und die stärkere Berücksichtigung theoretischer Erkenntnisse bei der empirischen Forschung andererseits werden im Teilprojekt A5 exemplarisch auf die Untersuchung von Polaritätselementen des Deutschen übertragen. Dieser Datenbereich setzt die Subregularitätsproblematik fort, da es verschiedene Gruppen von Polaritätselementen gibt und darüber hinaus individuelle Distributionsbeschränkungen vorliegen können. Auf zentralen theoretischen Forschungsergebnissen im Bereich der Polaritätselemente aufbauend sollen empirischen Studien durchgeführt werden, um so zu einer solideren Erfassung des Datenbereichs zu gelangen und eine Eva-

A



luationsbasis für die Theorien zu schaffen. Hierbei ergeben sich durch die korpusbasierten Untersuchungen und durch die systematische Erhebung von Sprecherurteilen Verbindungen zu A1 sowie zu A3 und A4. Da der Datenbereich in logischer Verlängerung der Arbeit der vorangegangenen Phase vor allem Fragen der semantischen Intuitionen, Repräsentationen und Kombinationsmechanismen betrifft, ist eine intensive Methoden- und Ergebnisdiskussion mit A4 anvisiert. Durch eine theoriegeleitete Implementierung soll eine Rückkopplung zwischen Empirie und Theorie entstehen.