

Kontextualisierte Wortschatzanalyse (KWA)

**Ziele, Probleme und Verfahren
ENDFAS Arbeitspapier Nr. 1 1992**

**Wilhelm Griebhaber
Jochen Rehbein**



sprachen ●●● zentrum

Westfälische Wilhelms-Universität

Münster – PALM 11/2002

© W. Griebhaber, J. Rehbein

Kontextualisierte Wortschatz-Analyse (KWA)

Ziele, Probleme und Verfahren

ENDFAS Arbeitspapier Nr. 1 • 1992

W. Griebhaber • J. Rehbein¹

Universität Hamburg • Germanisches Seminar

Inhaltsverzeichnis:

- 1 Ziel der KWA
- 2 Wortschatzanalyse-Projekte
- 3 Wortklassen, Wortarten
- 4 Bestimmung von 'Wort' / Symbolfeldausdrücken
- 5 Kontext und der neue Zugang zur Wortschatzanalyse
 - Einzelne Forschungsergebnisse
 - Kontext und Mittel des Symbolfeldes
 - Anforderungen an eine kontextsensitive Datenerfassung
- 6 Aufbereitung der Daten, Lemmatisierung
- 7 Modular getrennte Deskription
- 8 Wortbezogene Deskriptoren
- 9 Einzelfragen der Lemmatisierung
 - Allgemeine Deskriptionsgrundsätze
 - Lemmatisierung von Verben
 - Inkorporierte Suffixe im Türkischen
 - Behandlung komplexer und diskontinuierlicher Ausdrücke
 - Wortübergreifende Ausdrücke: türkische izafet-Bildungen
 - Wortübergreifende Ausdrücke: Zusammengesetzte Verweis- und Fragewörter
 - Wortübergreifende Ausdrücke: feste Verbindungen mit Verben
- 10 Semantische Deskription
- 11 Auswertung

Anhang:

Beispiele für Datenrecherchen

Schematische Darstellung der Analyseschritte

Semantische Deskriptoren (Neuner & Schade 1979)

Textbeispiele deutsch - türkisch

Literaturverzeichnis

¹ Grundlegende Ideen des Ansatzes wurden mit I. Karasu diskutiert, dem wir an dieser Stelle danken.

§1 Ziel der KWA

Mit der KWA sollen Einblicke in Umfang und Struktur des Wortschatzes in Abhängigkeit vom Verwendungskontext gewonnen werden. Untersuchungen des Wortschatzes jugendlicher oder erwachsener Sprecher sehen sich mit dem Problem konfrontiert, eine offene Menge zu bestimmen: einerseits ist der Wortschatz einer Sprache nicht abgeschlossen, sondern in fortwährender Entwicklung, andererseits ist weder in einer einzigen noch in einer Folge von Verwendungskonstellationen der gesamte Umfang des verfügbaren Wortschatzes einer Person aktivierbar, da sich der individuelle Wortschatz in ständiger Entwicklung befindet und in Abhängigkeit von verschiedenen Faktoren unterschiedlich zugänglich ist.

Untersuchungen sehr früher Spracherwerbsstadien erlauben in Longitudinalstudien die Erhebung und Auswertung weitgehend geschlossener Korpora. Spätere Erwerbsstadien dagegen bieten prinzipiell nur die Erhebung einzelner diskreter Korpora. Die Standardisierung von Erhebungskontexten und -aufgaben gestattet bei einer reduzierten Natürlichkeit des Verwendungszusammenhangs die Erhebung vergleichbarer Daten, die eine Einstufung der Probanden im Hinblick auf die Aufgabenstellung relativ zu anderen Vergleichsgruppen oder -personen zulassen. Insgesamt sind die Aussagen zum Wortschatz jedoch prinzipiell begrenzt.

Ausgehend von einer Bestimmung des **Untersuchungsgegenstandes** "Wortschatzanalyse" ist zunächst der Gegenstand der KWA genauer zu definieren. Insbesondere ist es notwendig, den Wortschatz nicht als 'totes' Lexikon zu betrachten. Stattdessen ist der individuelle Wortschatz im Zusammenhang mit dem Erwerbs- und Verwendungskontext der verschiedenen sprachlichen Mittel zu sehen. Damit ist eine Verbindung zum sozial-biographischen Hintergrund gegeben. Weiterhin kommt den Einheiten des Wortschatzes ein unterschiedlicher funktionaler Status bei der Prozessierung von Wissen beim sprachlichen Handeln zu: die Elemente des Symbolfeldes (im Bühlerschen Sinn) repräsentieren vielfältig gegliederte und strukturierte Einheiten des Wissens, und sind durch semantische und syntaktische Beziehungen zwischen den Einheiten miteinander verbunden.

Die **Wortschatzerhebung** sollte demnach mehrere Verwendungskonstellationen umfassen, insbesondere monologische und dialogische. Weiterhin sollte die Aufgabenstellung für die Verbalisierung variieren und zumindest folgende Kontexte umfassen: freies Wiedergeben fiktiver und/oder selbsterlebter Begebenheiten, gebundene Versprachlichungen bildlicher Vorlagen. Ergänzungen durch thematisch zentrierte assoziative Verbalisierungen oder Übersetzungsaufgaben sind möglich. Auf jeden Fall sollten mündliche Äußerungen den wesentlichen Teil der Daten ausmachen.

Bei Analysen bilingualer Kinder sollte die Elizitierung mutter- und zweitsprachlicher Äußerungen nicht von ein- und demselben Interviewer durchgeführt werden. Kinder

beziehen die verwendete Sprache auf den Gesprächspartner. Dementsprechend können sich systematische Verzerrungen ergeben, wenn beide Themen von einer Person bearbeitet werden, die für das Kind mit einer bestimmten Sprache verbunden ist: für das Kind dient diese personengebundene Sprache als Mittel der Informationsvermittlung, während die 'zweite' Sprache, die diese Person auch beherrscht, in weit stärkerem Maße als Mittel zur Prozessierung von Inhalten betrachtet wird.

Auch die **Aufbereitung** und **Analyse** der erhobenen Daten muß die oben vorgestellten Kontextbezüge berücksichtigen. Die übliche Speicherung von Äußerungen mit anschließender Lemmatisierung erlaubt zwar die Darstellung der aufbereiteten Daten in der Form von Matrizen nach Wortformen, Wortarten, semantischen Feldern usw.; sie bietet jedoch kaum Einblicke in die Strukturiertheit und Systematik des Wortschatzes von Probanden. Deshalb ist ein Verfahren erforderlich, das systematisch die Verwendungskontexte berücksichtigt und Bezüge zwischen folgenden Größen zuläßt: (1) den sozial-biographischen Kontexten der Probanden, (2) den äußerungsexternen Erhebungskontexten (Art der Äußerungsevozierung, Verwendungszusammenhang, ...) und (3) den äußerungsinternen Verwendungskontexten.

§2 Wortschatzanalyse-Projekte

Ist schon die Zahl der Untersuchungen zum zweitsprachlichen Wortschatzerwerb gering (s. Meara (1983) und (1987)), so sind Analysen zur zweisprachigen Wortschatzentwicklung äußerst spärlich. Über lange Zeit hinweg standen in der Fremdsprachendidaktik grammatische Strukturen im Mittelpunkt. Lado (1969²) formuliert den Vorrang der Strukturen gegenüber der Lexik folgendermaßen: "Während der Einübung des Lautsystems und der grammatischen Strukturmuster ist die Wortschatzerweiterung auf ein Minimum zu reduzieren." (S. 79) und: "Im Mittelpunkt des Unterrichts stehen Sätze, nicht Wörter" (S. 81). Auch in der Zweitspracherwerbsforschung dominierten Untersuchungen zur Morphologie und Syntax, sowohl in eher strukturalistischen wie in Arbeiten im Rahmen der generativen Transformationsgrammatik, die mit dem Anspruch der Beschreibung der zugrundeliegenden Sprachkompetenz eines Sprechers/Hörers sich lange auf die Syntax konzentrierte. auf der anderen Seite kommen sowohl Studien zum Erstspracherwerb wie zum gleichzeitigen Erwerb zweier Sprachen zu dem Ergebnis, daß der Wortschatzerwerb in starkem Maße von kontextuellen Faktoren bestimmt wird.

Für den Erstspracherwerb hat im psycholinguistischen Ansatz von Clark & Clark (1977) Clark, E. (1979) ein Konzept zur Ausdifferenzierung semantischer Komplexität im Hinblick auf die semantische Rollen und Eigenschaftskomplexe vorgelegt. Diese Studie berücksichtigt besonders den sehr frühen Spracherwerb, der im vorliegenden Arbeitszusammenhang nicht im Zentrum steht. Als wichtige Steuerungsfaktoren bestimmt sie folgende Kontextfaktoren: die an die Kinder adaptierte Sprache der Erwachsenen, das wachsende Wissen der Kinder im Verlaufe ihres Entwicklungsprozesses und die inhärente Komplexität von Wortbedeutungen.

Taeschner (1983) untersucht in ihrer Studie zum zweisprachigen Spracherwerb die Entwicklung des

Wortschatzerwerbs (Kapitel 2). Auch sie konzentriert sich auf den hier ausgeklammerten Bereich des sehr frühen Spracherwerbs von den ersten Äußerungen bis etwa 3 Jahre. Dabei unterscheidet sie zwei Stadien. Im ersten benutzt das Kind Wörter aus beiden Sprachen, während im zweiten die Worte nach Sprachen getrennt werden, indem Äquivalente erworben werden. Dabei betont sie die Bedeutung pragmatischer, d. h. kontextueller, Faktoren für die Entwicklung des Lexikons.

Im folgenden sollen kurz zwei Untersuchungen zum Wortschatzerwerb vorgestellt werden, um die Fragestellungen und Probleme eines solchen Unternehmens herauszuarbeiten. Die eher traditionelle Arbeit von Hepsöyler & Liebe-Harkort (1988), die den Wortschatz bilingualer türkischer Schüler in der Bundesrepublik mit dem Wortschatz monolingualer türkischer Schülern in der Türkei, bzw. deutscher Schüler in der Bundesrepublik vergleicht, wird nicht ausführlich vorgestellt. Die Untersuchung von Mrazovič & Stölting-Richert (1989) zur Balance des Wortschatzes von bilingualen serbokroatischen Schülern in der Bundesrepublik wird ausführlicher vorgestellt. Außerdem wird eine Studie zum Wortschatzerwerb von Erwachsenen in den Niederlanden besprochen, die sich nur mit dem Erwerb des Niederländischen befaßt, jedoch eine chronologische Entwicklung beschreibt und ein umfangreiches quantitatives Instrumentarium einsetzt.

(1) Mrazovič & Stölting (1989)

Ziel der Untersuchung ist die Klärung des Verhältnisses zwischen den deutschen und serbokroatischen Wortschatzen von jugoslawischen Schülern mit serbokroatischer Muttersprache, die Deutsch als Zweitsprache in Deutschland erwerben. Die Datenerhebung erfolgte im schulischen Kontext und umfaßte grundsätzlich monologische Äußerungen der Probanden im jeweils monolingualen Modus. Durch eine breite Palette von sieben schriftlichen und einer mündlichen Aufgabe sollten unterschiedliche Aspekte der Wortschatzbeherrschung erfaßt werden.

Die schriftlichen Aufgaben:

- (1) *Cloze Test*: Lücke jedes 9. Wort;
- (2) *Wörternennen*: zu zwei schriftlichen Stimuli (dt: Geburtstag/Osterferien; skr: rođendan/uskrsnji praznici) in je 4 Minuten weitere Wörter (Substantive, Adjektive, Verben, Eigennamen) assoziieren; Ziel: Wortverfügbarkeit, Grad der Übersetzungsäquivalenz;
- (3) *Aufsatz*: "Mein allerschlimmstes Erlebnis" / "Moj najneprijatniji doživljaj"
- (4) *Phraseologie-Test*: schwierige idiomatische Redewendungen; je Sprache 8 Aufgaben; Erklärung der Redewendung, idiomatische Wiedergabe in der anderen Sprache;
- (5) *Anagramm-Test*: Bildung neuer Wörter aus den Buchstaben vorgegebener Wörter; Ziel: Wortverfügbarkeit, orthographisch korrekte Wörter;
- (6) "Wir suchen Reime" Reimbildung aus isolierten Stimulus-Wörtern; 10 Min. pro Sprache;
- (7) "Reim dich oder ich fress dich" / "Slaze se kao sou oci" Ausfüllung von Wortlücken in einem vorgegebenen Originalgedichte unter Berücksichtigung von Reim und Rhythmus.

Die mündliche Aufgabe:

Filmmacherzählung: Stummfilm über einen Birnendiebstahl.

Um die beiden Wortschatze zueinander in Beziehung setzen zu können, wurden quantitativ erfaßbare Daten in Diagrammen so angeordnet, daß die sprachbezogenen Schwerpunkte deutlich sichtbar sind. Qualitative Aussagen, z. B. zur Ausdrucksdifferenzierung, erfolgen in verbalen Beschreibungen. Mit zunehmender Schulzeit wächst im allgemeinen der elizitierte Wortschatz in beiden Sprachen.

(2) ESF-Projekt (Broeder & Extra & van Houten (1989))

In der Studie, die im Rahmen des europäischen Forschungsprojekts der ESF zum Zweitspracherwerb von Erwachsenen durchgeführt wurde, wird untersucht, wie vier Erwachsene Niederländisch erwerben. Zwei der Sprecher haben Arabisch als Muttersprache, zwei Türkisch. Grundlage sind mündliche Äußerungen, die in drei zeitlich in großen Abständen durchgeführten Aufnahmesitzungen erhoben wurden. Sie umfassen jeweils ein freies Gespräch mit einem niederländischen Muttersprachler und die Wiedergabe eines (verschiedenen) Stummfilms.

In einer quantitativen Makro-Perspektive werden Umfang und Variabilität des Lexikons in Abhängigkeit von den Aufgaben und den Aufnahmen untersucht. In qualitativ orientierten Mikroanalysen werden semantische Domänen nominaler und pronominaler Referenz untersucht. Die Daten werden so aufbereitet, daß Zähloperationen durchgeführt werden können. Im einzelnen werden die Daten in folgendem Verfahren für die Durchführung der Analyse aufbereitet:

- (1) *Erstellung von Konkordanzlisten:* die von den Lernern benutzten Wortformen werden in alphabetischer Reihenfolge zusammen mit dem Kontext und der Anzahl aufgelistet;
- (2) *Erstellung von Worttyplisten:* die Konkordanzlisten werden gesäubert (z. B. von "false starts") und in Worttyplisten überführt; jeweils identische "word tokens" bilden eine "word form";
- (3) *Überführung der Worttyplisten in eine Datenbank:* jede Wortform wird in einem Datensatz ("record") gespeichert, der neben der Wortform noch die grammatisch bestimmte Wortklasse, die vermutete Wortbedeutung, das Lemma, die Häufigkeit und Referenzangaben enthält (Informant, Aufnahmezyklus, Begegnung und Aktivität).

Umfang und Variabilität werden neben globalen Type / Token Relationen mit berechneten Indizes ermittelt. Verwendet werden Girauds Index ($\text{number of types} / \sqrt{\text{number of tokens}}$) und eine Größe zur Variabilität des Lexikons bezogen auf eine normierte Anzahl von Wörtern. Insgesamt zeigt sich, daß Umfang und Variabilität in den freien Gesprächen über den Werten für die Filmmacherzählungen liegen. Über die Zeit hinweg läßt sich ein zunehmender Umfang feststellen.

Aus der Beschreibung des ESF Projekts geht nicht eindeutig hervor, ob die Wörter von den von den Lernern verwendeten Vollformen auf lemmatisierte Grundformen zurückgeführt und ausgewertet werden, oder ob gleiche Vollformen innerhalb einer Wortart als Grundeinheit verwendet werden. Sicher scheint, daß alle Wörter, also auch Funktionswörter oder Interjektionen, in die Makroanalyse eingehen. Da die untersuchten Einheiten unabhängig von ihrer Funktion gemeinsam untersucht werden, sind die quantitativ ermittelten Ergebnisse wenig aussagekräftig im Hinblick auf die Frage, über welchen Wortschatz ein Sprecher verfügt. So weist die Sprecherin Fatima in allen sechs Verwendungskontexten sowohl nach Types wie nach tokens die geringsten Zahlen auf, doch ist dies beim theoretischen Vokabular nicht mehr eindeutig der Fall und bei Girauds Variabilitätsindex belegt sie auch nicht immer den letzten Platz. Dennoch weist sie mit insgesamt 329 belegten pronominalen Wortformen nur rund ein Drittel des Wertes auf, den der nächstniedrige Sprecher mit 1074 aufweist. Fatima muß demnach erheblich mehr 'Inhaltswörter' verwendet haben als die übrigen drei Sprecher, ohne daß dies in der quantitativen Makroanalyse sichtbar wird.

§3 Wortklassen, Wortarten

Die Aussagekraft von Wortschatzanalysen hängt wesentlich davon ab, wie der Ge-

genstand der Untersuchung bestimmt wird. Vor allem sollten Einheiten mit eigener lexikalischer Bedeutung erfaßt und untersucht werden. Sog. funktionale Strukturwörter - Konjunktionen, Artikel, Präpositionen, Pronomina - werden zwar häufig verwendet, lassen jedoch kaum Rückschlüsse auf die Breite und Tiefe des Wortschatzes zu, über den ein Sprecher verfügt.

In diesem Abschnitt sollen exemplarisch einige Aspekte der Bestimmung von Wortarten diskutiert werden. Der folgende Überblick zeigt, daß sich der Wortschatz nach sehr verschiedenen Kriterien erfassen und strukturieren läßt. Doch auch eine klare Bestimmung der zu untersuchenden Einheiten ist noch nicht ausreichend. Ossner (1989) demonstriert an einem deutschen Beispielsatz², daß fünf neuere spezielle Arbeiten zu den Wortarten und Grammatiken des Deutschen (Bergenholtz & Schaefer (1977), Engel (1988), Helbig (1984), Duden (1984) und die Grundzüge (1981)) zu unterschiedlichen Ergebnissen gelangen. Beispiele aus der (frühen) Maschinellen Sprachverarbeitung zeigen, daß auch klare Bestimmungen nicht einfach zu operationalisieren sind.

Traditionell erfolgt die Wortartenbestimmung vor allem nach folgenden drei Gliederungsaspekten (s. z. B. Engel (1988: 17f.)):

- (a) **Morphologie:** bei den flektierenden europäischen Sprachen bietet sich der Aspekt der morphologischen (Un-)Veränderbarkeit der Einheiten des Lexikons als Unterscheidungskriterium an, da sich die Äußerungen damit relativ klar in distinkte Klassen einteilen lassen: in die flektierenden Wortarten einerseits (Substantiv, Adjektiv, Verb, Pronomen) und die nicht flektierenden Wortarten andererseits (Adverb, Konjunktion, Präposition);
- (b) **Syntax:** unter syntaktischen Gesichtspunkten lassen sich Wörter nach ihrer Fähigkeit gliedern, als Satzglied verwendbar zu sein, nominale oder verbale Elemente zu modifizieren, einen Artikel zu sich zu nehmen (Substantiv vs. Pronomen), oder mittels Rektion einen bestimmten Kasus zu fordern (Präposition vs. Konjunktion);
- (c) **Semantik, bzw. begriffliches Kategoriensystem:** schon antike Sprachforscher legten die Grundlagen für eine semantisch, bzw. begrifflich-kategorial bestimmte Definition des Lexikons in die drei Grundwortarten Substantiv, Adjektiv und Verb, denen die logischen Kategorien 'Substanz', 'Eigenschaft' und 'Prozeß' entsprechen, während der Konjunktion und Präposition die Kategorie der 'Relation' entspricht.

Die zuletzt genannten Systeme stehen in einer alten philosophisch-sprachwissenschaftlichen Tradition, erfassen in ihren reinen Ausprägungen jedoch nicht die funktionale Leistung der verwendeten Mittel in den konkreten Äußerungen. Die ersten beiden Vorgehensweisen orientieren sich in ihren reinen Formen an äußerlich wahrnehmbaren Merkmalen und teilen die Wörter nach diesen Merkmalen in Klassen ein. Dadurch weisen sie inhärente Klassifikationsprobleme auf, insofern das jeweils

² "Der Grammatiker versucht jede Wortklasse eigentlich völlig neu zu bestimmen." (S. 108)

zugrundegelgte rein formale Kriterium nicht nur zu Überschneidungen mit anderen Kriterien, sondern auch zu anderen Klasseneinteilungen führt. Bei einer morphologisch basierten Einteilung enthält die Gruppe der flektierenden Wortklassen mit den Adjektiven und Substantiven den größten Teil der Elemente, die intuitiv bei einer Wortschatzanalyse erfaßt werden sollten. Sie enthält jedoch darüberhinaus mit den Pronomina auch eine Wortklasse ohne eigene Bedeutung, die zudem in sich selbst höchst heterogen ist. Die Gruppe der nichtflektierenden enthält mit den Adverbien eine Wortklasse, die intuitiv unbedingt bei der Analyse erfaßt werden sollte. Ähnliche Gruppierungsprobleme ergeben sich auch bei einer syntaktisch basierten Einteilung, da das syntaktische Bindungspotential nicht mit dem Kriterium der Eigensemantik einhergeht, wie das Beispiel der Präpositionen mit ihrer geringen Autosemantik bei ausgesprochener Rektionskraft zeigt.

Die Saarbrücker Forschungsgruppe orientierte sich bei der Bestimmung von Einträgen der Klassenzugehörigkeit von Wörtern zunächst an traditionellen Grammatiken (u. a. dem DUDEN), um dann folgende, nicht untypische Probleme zu beschreiben:

"Es zeigte sich, daß wir für unser maschinelles Analyseverfahren eine Reihe von Modifikationen vornehmen mußten. Dies trifft besonders zu für die mehr inhaltlich orientierten Definitionen der 'Wortarten'. Prinzipiell dürften nach der Methode strenger Distribution nur diejenigen Wortformen in einer Klasse zusammengefaßt werden, die in allen Sätzen gegeneinander austauschbar sind, ohne daß sich dabei die übrige Wortformenfolge ändert. Es war uns nicht möglich, unser Analyseverfahren auf derartig definierten Wortklassen aufzubauen, ... Wir mußten uns daher für einen Mittelweg entscheiden, da auch die Wortartenaufteilung der Grammatiken nicht ohne Änderung übernommen werden konnte. ... Allerdings veranlaßt uns die Notwendigkeit einer rein formalen Klassifizierung, Wortklassen zu bilden, die von der herkömmlichen Einteilung abweichen. Dadurch bekommen zum Teil auch die herkömmlichen Bezeichnungen - soweit wir sie verwenden - einen anderen Inhalt. So gelten uns z. B. als 'Adjektiv' (ADJ) nur die attributiven, meist flektierten Adjektiva, aber auch die entsprechenden Partizipia; zum Adverb rechnen wir außer den herkömmlichen Gliedern dieser Gruppe auch die unflektierten Adjektiva und die Partizipia, soweit sie nicht als Wortklasse 'Partizip II' (PTZ) als Teil einer Verbalgruppe zu betrachten sind." (Zimmermann (1969:57).

Das ebenfalls auf die maschinelle Sprachverarbeitung zurückgehende syntaktisch basierte Klassifikationsmodell von Bergenholtz & Schaefer (1977) kommt auf 51 Wortklassen. Die hohe Zahl resultiert aus der strikten Oberflächenorientierung, die Stellungsvarianten mit eigenen Wortklassen berücksichtigt. Beispielsweise umfaßt die Klasse der Präpositionen üblicherweise nicht nur die vorangestellten Präpositionen, sondern aufgrund gleicher funktionaler Leistung auch nachgestellte 'Präpositionen' (Postpositionen) oder Kombinationen aus vor- und nachgestellten 'Präpositionen' (Circumpositionen). Bergenholtz & Schaefer behandeln diese Varianten jedoch als eigene Wortklassen.

Der Exkurs in die Operationalisierung von Klasseneinteilungen zeigt einige der Probleme, die zu lösen sind, wenn authentische Äußerungen nach einem einheitlichen Set von Kriterien eindeutig bestimmt werden sollen. Traditionelle Arbeiten nehmen in nicht explizit ausgewiesener Art die Fähigkeit des kompetenten menschlichen

Sprechers/Hörers in Anspruch, um Mischklassifikationen beim Vollzug der Zuordnung kreativ auf das vorfindbare Material anzuwenden. Dabei ist es möglich, Wörter, die den vorgegebenen Kriterien nicht eindeutig entsprechen, nach übergeordneten funktionalen Kriterien in eine der vorgegebenen Gruppen einzuordnen. Die aktuell vorfindbare Stellung von 'Präpositionen' in einem Korpus spielt dabei eine untergeordnete Rolle. Der Versuch, solche Sortierarbeiten von Computern durchführen zu lassen, zeigte die Lücken eines solchen Vorgehens.³ In diesem Sinne kommt der Nutzung maschineller Sprachverarbeitungssysteme eine heuristische Funktion zu, da der Zwang zu einem strikt formalisierten Vorgehen zu wesentlich genaueren und expliziteren Operationalisierungen führt.

Solche Probleme sind auch beim Verfahren der KWA zu berücksichtigen. Die Überführung der Äußerungen der Probanden in strukturierte Wortschatzlisten und die computergestützte Auswertung dieser Listen erfordern ein operationalisiertes Verfahren für die Erstellung der Wortschatzlisten und die Zuweisung der Wörter zu einer Wortart. Allerdings sind aufgrund der oben gezeigten Probleme mit formalen, Syntax und/oder Morphologie basierten Klassifikationsverfahren für eine Wortschatzanalyse wenig geeignet. Zum einen lassen sich gerade mit formalen Verfahren die vom Untersuchungsziel her notwendigen Differenzierungen zwischen Wörtern mit eigener Bedeutung und anderen Wörtern schlecht oder nicht treffen. Zum anderen führt der Vorschlag von Bergenholtz & Schaefer drastisch vor Augen, daß ein strikt formales Vorgehen eine kaum zu bewältigende Fülle von Deskriptoren bedingt. Die Wortschatzanalyse erfordert also ein Vorgehen, das von der funktionalen Bestimmung der Wörter ausgeht und entsprechende Kriterien für die Zuordnung der in den Äußerungen auffindbaren Wörter zu den Wortarten bereitstellt. Gleichzeitig ist von vornherein zu akzeptieren, daß die Bestimmung der Wortarten eine aktive Tätigkeit ist, die sich nur partiell formalisieren läßt, und deshalb weitgehend manuell durchgeführt werden muß.

³ Gerade auch das von Engel (1988: 20) vorgestellte Diagramm zur Wortklassenbestimmung nimmt nicht offengelegten Klassifikationsprozeduren in Anspruch. So werden die Präpositionen in einem Zweig behandelt, in dem die zu klassifizierenden Elemente zuvor als unveränderliche bestimmt wurden. Wenn man nach diesem Kriterium auch solche Formen wie "beim", "am", "durchs" usw. als Präposition erfassen will, muß man diese Variationen der Form als besondere Veränderung erkennen, die sich von der Flexion von Verben oder Substantiven unterscheidet.

§4 Bestimmung von 'Wort' / Symbolfeldausdrücken

Ein funktionaler Zugang zur Wortschatzanalyse muß bei der Wortartenbestimmung von der Leistung der Wörter bei der Verwendung ausgehen. Einen geeigneten Ausgangspunkt bietet das semiotisch-funktionale Vorgehen von Bühler (1934). Er bezieht sprachliches Handeln systematisch auf die Grundkonstellation von Sprechern und Hörern, die sich unter Bezug auf eine Origo im Hier und Jetzt miteinander sprechend über präsente und absente Sachverhalte verständigen. Dabei entwickelt er das Konzept zweier sprachlicher Felder, das Zeigfeld und das Symbolfeld, denen jeweils entsprechende Wortarten zugeordnet sind, dem Zeigfeld bsp. die Zeigwörter, dem Symbolfeld die Nennwörter.

Mit den Mitteln des Zeigfeldes erreicht der Sprecher eine Orientierung des Hörers im Sprechzeitraum, er steuert ihn unter Verwendung deiktischer Ausdrücke mittels der deiktischen Prozedur. Bedeutung im engeren Sinn haben die Mittel des Symbolfeldes. Worte besitzen nach Bühler eine phonematische Prägung und Feldfähigkeit (§19). Sie besitzen einen 'Sinn' und haben eine Funktion im Sprechverkehr. Ihre Kombinationsfähigkeit mit anderen Wörtern im Satz werden durch Feldmarken gesteuert, die ihre Verbindungsfähigkeiten bestimmen. Worte haben demnach eine akustische Gestalt, die sie individualisiert, von anderen Worten abgrenzt und formale Kennzeichen, durch die ihre Zugehörigkeit zu einem der beiden Felder ausgedrückt ist und durch die ihre Kombinationsfähigkeit mit anderen Wörtern bestimmt wird.

Ehlich (1979) interpretiert die von Bühler sprachpsychologisch entwickelten Felder um, indem er sie auf das sprachliche Handeln bezieht und ihnen jeweils spezifische Prozeduren zuordnet. In einer späteren Arbeit erweitert er Bühlers Zweifeldermodell auf insgesamt fünf Felder (Ehlich (1986)). Den von Bühler beschriebenen Zeig- und Symbolfeldern fügt er das Malfeld, das Lenkfeld und das Operationsfeld hinzu. Die für die Wortschatzanalyse relevanten sprachlichen Ausdrücke sind dabei dem Symbolfeld zuzurechnen, mit dessen Mitteln die bezeichnende Prozedur vollzogen wird.

Die Mittel des Symbolfeldes sind in mehrfacher Hinsicht strukturiert. In Abhängigkeit von der Zugehörigkeit zu einer bestimmten Wortart besitzen sie eine spezifische Potenz zur Organisation des propositionalen Gehalts in einer Äußerung. Verben steuern über die Verbflexion die Verankerung des propositionalen Gehalts in Bezug auf den Sprechzeitraum (s. Rehbein (1991)) und eröffnen ein jeweils spezifisches Feld für Beziehungen zu nominalen Elementen. Den Kern der Symbolfeldausdrücke bilden im Deutschen Eigennamen, Substantive, Adjektive, Adverbien und Verben.

Für die Untersuchung des Wortschatzes im Deutschen und Türkischen sind an dieser Stelle einige prinzipielle Fragen zu klären. Das zur Familie der indoeuropäischen Sprachen gehörende Deutsche ist eine flektierende Sprache. Das bedeutet in diesem Zusammenhang, daß einzelne lexikalische Elemente zumindest etymologisch einer

Wortart zugewiesen werden können. Entsprechend dieser Zugehörigkeit tragen die Angehörigen einer Wortart charakteristische Merkmale (Bühlers "Feldmarken"). Substantive besitzen im Deutschen ein Genus, das zwar nicht direkt am Wort sichtbar sein muß, aber in Verbindungen mit Artikeln oder Adjektiven an deren Flexionsendungen deutlich wird. In Abhängigkeit von der Wortart können die Wörter bestimmte Kombinationen mit anderen Wörtern eingehen und bestimmte Flexionssuffixe annehmen. Dabei sind im allgemeinen verschiedene Funktionen in einem komplexen Suffix verschmolzen.

Das Türkische dagegen ist eine agglutinierende Sprache, bei der die einzelnen Funktionen in eigenen Suffixen repräsentiert sind, die als solche in einer bestimmten Reihenfolge an den Stamm angefügt werden. Bedeutsamer für die Wortschatzanalyse ist jedoch die Tatsache, daß den Wörtern im Türkischen klare Eigenschaften fehlen, die sie vom Stamm her fest einer Wortart zuweisen. Je nach äußerungsinternem Kontext und angehängten Suffixen kann ein Stamm adjektivisch, substantivisch oder verbal benutzt werden (s. dazu den Überblick von Johanson 1990, Abschnitt 1). Die Bestimmung richtet sich deshalb nach der aktuellen Verwendung. Vor diesem Hintergrund können Wörter, die in den beiden Sprachen den gleichen Wortarten zugewiesen werden, nicht direkt gleichgesetzt werden (s. u.).

Umfang, Strukturiertheit der Symbolfeldausdrücke

Die Elemente des Symbolfeldes sind mehrfach strukturiert. Bühler unterscheidet die Sphäre, der ein Symbolfeldausdruck zuzurechnen ist und das Etymon, d. h. den bei der ursprünglichen Benennung hervorstechenden Inhalt. Letzterer kann verblässen oder verschwinden und muß synchronisch den Ausdruck nicht mehr regieren. Mit der 'Sphäre' ist der Umfang eines 'Bedeutungsfeldes' gemeint. Schließlich können sich einzelne Sprecher darin unterscheiden, wie sie ihren Wortschatz mit Mitteln der Wortbildung ausdifferenzieren. Die Beschreibung der Wortschatzentwicklung von zweisprachig aufwachsenden Kindern soll die kontextualisierte Wortschatzanalyse u. a. zur Beantwortung folgender Fragen beitragen (s. auch Bühlig 1992 zum Situations-sampling):

- (a) Unterscheiden sich die Umfänge in beiden Sprachen? Wenn ja, lassen sich die Unterschiede mit unterschiedlichen, möglicherweise lebensweltlich mit einer Sprache dominanten Verwendungsweisen, oder Bezugspersonen in Verbindung bringen?
- (b) Gibt es sprachimmanente Unterschiede in der Strukturiertheit des Wortschatzes?
- (c) Welche Unterschiede treten in Abhängigkeit vom Erhebungskontext und der gestellten Aufgabe auf?
- (d) Lassen sich Unterschiede im Wortschatz mit unterschiedlichen Erwerbsverläufen, Unterrichtsmethoden usw. in Verbindung bringen?
- (e) Welche Zusammenhänge gibt es zwischen Umfang und Strukturiertheit des Wortschatzes und dem Erzählen?

- (f) Gibt es Zusammenhänge zwischen den verwendeten Wortbildungsmitteln und dem Erzählen?

Zur Beantwortung solcher Fragen erweisen sich vorliegende semantische Klassifikations- und Beschreibungssysteme aufgrund ihres jeweiligen Verwendungszweckes als wenig geeignet. Dornseiff (...) verwendet ein sehr detailliertes System, das sämtliche Wissensbereiche einbezieht. Seine Anwendung käme bei der vorliegenden Untersuchung einem Overkill gleich ohne angesichts der Abstraktheit der Deskriptoren eine angemessene Differenzierung der zu untersuchenden Kinder zu gewährleisten. Ähnliche Probleme ergeben sich bei linguistischen Wortschatzanalysen. Sie gehen im allgemeinen sprachsystematisch vor und zielen auf eine möglichst differenzierte Beschreibung eines abgegrenzten lexikalischen Bereichs ab, die die Grundlage für anschließende Verallgemeinerungen bildet. Dies zeigt sich an den Untersuchungen zu Verben allgemein (s. Ballmer & Brennenstuhl 1986) oder zu ausgewählten Verbgruppen (z. B. Fortbewegungsverben, Diersch 1972, die Triers (1931) Wortfeldansatz weiterführt). Diese Arbeiten sind abhängig von den zugrundegelegten Bedeutungsbegriff und den dafür gewählten Unterscheidungskriterien, die für den vorliegenden Untersuchungszweck kaum übertragbar sind. Die Klassifikation der Zustands- und Bewegungsverben von Gerling & Orthen 1979 kommt zu derart feinen Differenzierungen nach der Art der Fortbewegung und den dabei verwendeten Hilfsmitteln, daß sie für eine praktische Anwendung in der Deskription ungeeignet ist.

Einen weiteren Ansatzpunkt bieten die von anderen Autoren bei der Wortschatzanalyse von Kindern verwendeten Deskriptoren. Diese Arbeiten gehen von bestimmten Korpora aus und entwickeln Kriterien, die zur Beschreibung des jeweiligen Korpus für die aktuellen Untersuchungszwecke geeignet erscheinen. Neuner & Schade (1986) gelangen so bei der Analyse von Lehrmaterialien zu einem Inventar von 18 Deskriptoren (s. Anhang). Diese Deskriptoren reflektieren in stärkerem Maß die Lebenswelt und den potentiellen Wortschatz von Kindern. Sie sind jedoch auf komplexe Tätigkeiten bzw. Tätigkeitskomplexe bezogen und nicht auf die semantische Beschreibung von Einzelwörtern. Dadurch können einzelne Wörter mehreren Bereichen zugeordnet werden. "Fernsehen" kann z. B. sowohl mit dem Deskriptor "Umwelt" wie "Freizeit / Unterhaltung" oder auch "persönl. Beziehungen" zugewiesen werden, wenn das Fernsehen von den Kindern beispielsweise als soziale Tätigkeit mit Freunden und Verwandten dargestellt wird. Die KWA muß also geeignete Deskriptoren entwickeln. Dabei wird grundsätzlich induktiv von den vorliegenden Daten ausgegangen werden.

Eine weitere Dimension der Strukturiertheit des Wortschatzes betrifft die Verwendung lexikalischer Einheiten im Zusammenhang mit dem Wissen des Sprechers und des Hörers. Diese Einbeziehung der Handlungskomponente ergibt eine weitere Dimension der Strukturiertheit des Wortschatzes bezüglich des Status des Wissens und seiner diskursiven Markierung.

§5 Kontext und der neue Zugang zur Wortschatzanalyse

Die KWA berücksichtigt, daß die Einheiten des Wortschatzes einerseits bestimmte Funktionen bei der Organisierung und Prozessierung des sprachlich gefaßten Wissens erfüllen und andererseits in Abhängigkeit von den kommunikativen Verwendungsbedingungen in unterschiedlichem Maße zugänglich sind und systematisch in bestimmten Verbindungen auftreten. Kontext bedeutet demnach, daß das Wissen und die Verfügbarkeit lexikalischer Einheiten strukturiert ist, und daß die Verbalisierung in strukturierter Weise erfolgt.

Im Hinblick auf die Ziele der Wortschatzanalyse können folgende kontextuelle Aspekte unterschieden werden:

- (0) der sozial-kulturelle Kontext der Sprachgemeinschaft;
- (1) der individuell sozial-biographische Kontext (Biographie, Lernverfahren) sowie der Lern- & Verwendungskontext;
- (2) der äußerungsexterne Erhebungskontext, d. h. der Verwendungszusammenhang der sprachlichen Mittel, der Typ der gestellten Aufgabe (freie Erzählung, Bildbeschreibung, Assoziation, ...), der Äußerungsmodus, der monologisch oder dialogisch sein kann;
- (3) der äußerungsinterne Verwendungskontext der lexikalischen Einheiten, d. h. ihre isolierte (z. B. bei Assoziationsaufgaben), freie oder in Sätzen kombinierte Verwendung.

Für die kontextualisierte Wortschatzanalyse ergeben sich daraus zwei zu berücksichtigende Aspekte: (1) die Elizitierung sollte einen möglichst natürlichem Verwendungszusammenhang gewährleisten und (2) die Analyse sollte unter Berücksichtigung des Verwendungszusammenhangs erfolgen.

Im folgenden werden einzelne Forschungsergebnisse vorgestellt, die die Kontextabhängigkeit von Verbalisierungen zeigen. Anschließend werden daraus Schlußfolgerungen für die Elizitierung im Zusammenhang mit Wortschatzuntersuchungen vorgestellt. Zum Schluß werden Anforderungen an eine Erfassung und Auswertung der erhobenen Daten präsentiert, die gewährleisten sollen, daß bei der Analyse die oben vorgestellten Dimensionen des Kontextes in die Untersuchung einbezogen werden können.

Einzelne Forschungsergebnisse:

Verschiedene *Assoziationstests* belegen, daß Sprecher sprach- und kulturspezifisch strukturierte Ergänzungen zu Vorgaben machen. Die Vorlage eines Wortes eröffnet jeweils einen bestimmten alters- und sprachabhängigen Zugang zu einer spezifischen Reaktion. Kinder äußern eher ein textuell verbundenes Wort, also z. B. auf den Stimulus *Hund* die Reaktion *bellen*, Erwachsene eher ein textuell anderes Wort, also

z. B. auf den Stimulus *Hund* die Reaktion *Katze* (s. Ickler (1987: 11)). Clark (1970) teilt die Assoziationsverfahren nach zugrundeliegenden Regeln in paradigmatische (1-4) und syntagmatische (5-6) ein (nach Ickler, ebd.):

- (1) Regel des minimalen Kontrastes: long - short, good - bad, boy - girl;
- (2) Regel der Tendenz von der markiertheit zur Unmarkiertheit (eher als umgekehrt): better- good, dogs - dog;
- (3) Regel der Tilgung und Hinzufügung von Merkmalen (Super- und Subordination): apple - fruit, fruit - apple, kill - die;
- (4) Regel der Erhaltung der Wortart.
- (5) Regel der Realisierung ausgewählter Merkmale: young - boy, girl, child, man, people;
- (6) Regel der Vervollständigung fester Verbindungen: cottage - cheese.

Ickler kritisiert die Verbindung von Assoziationstypen mit unterschiedlichen psychischen Prozessen. Da die Experimente mit isolierten Wörtern vorgenommen wurden, wirkt das Fehlen des sonst im normalen Handeln gegebenen Kontexts seinerseits als Fokussierung, der unter Verwendung unwillkürlicher Mechanismen die üblichste Fortsetzung des begonnen 'Textes'. Da der übliche Hintergrund fehlt, wird oft ein Kontrastwort gewählt, um es als Alternative gegen das vorgegebene Wort abzusetzen.

Ehlich & Rehbein (1977) entwickeln bei der Analyse der Institution Schule Beziehungen zwischen der Institution und institutionsspezifischem Wissen. Sie unterscheiden verschiedene *Wissensstrukturentypen*. Wissen wird bestimmt als dreistellige Relation zwischen dem Subjekt des Wissenden (S), dem Thema des Wissens (Θ) und dem Gewußten (Γ). Die Strukturtypen unterscheiden sich im Hinblick auf ihre pragmatische Quantifizierung. Abgerundetes und fertiges Wissen über einen Gegenstandsbereich werden zu einem Bild geformt. In schriftlichen Assoziationsaufgaben zum Thema "Schule" finden sie Verarbeitungsprozesse, die aus mehreren Einschätzungen zum Thema Schule schließlich zum Bild führen "SCHULE IST SCHEISSE" (S. 51).

Im Zusammenhang der hier vorliegenden Leitfrage nach Kontextabhängigkeit von Wissensäußerungen kann daraus der Schluß gezogen werden, daß Assoziationsaufgaben Wissen über Gegenstandsbereiche mobilisieren und systematisierte und synthetisierte Wissenspartikel stimulieren, indem sie dafür einen Ausgangs- und Bezugspunkt setzen. Diese Aufgaben elizitieren insofern individuell basiertes Erlebnisswissen, das jedoch über die gesellschaftliche Einbindung des Individuums gesellschaftlich geprägt ist. Sie können - unter Berücksichtigung der Beschränkungen durch den jeweiligen Erhebungskontext - Aufschlüsse über den sozial-biographischen Hintergrund des Sprechers geben.

Rehbein (1982) beschreibt *begriffliche Prozeduren* türkischer Kinder bei der angeforderten Wiedergabe einer zuvor gesehenen Fernsehsequenz. Er untersucht die lexikalischen Strategien, die sie verwenden, um das Gesehene in der Zweitsprache Deutsch wiederzugeben. Aufgrund der vorherigen Filmrezeption kann davon ausgegangen

werden, daß sich die von den Sprechern verwendeten Prozeduren bei der Füllung lexikalischer Lücken auf ein vorhandenes Wissen beziehen. Die Probleme liegen in der Überführung dieses Wissens in propositionale Gehalte und deren Verbalisierung. Griebhaber (1987) analysiert *Sprecherstrategien* in authentischen und simulierten Einstellungsgesprächen von türkischen Bewerbern und deutschen Sachbearbeiterinnen. In Weiterführung von Rehbein (1982) werden lexikalische Strategien bei der Suche nach einem passenden Wort zur Bezeichnung eines spezifischen Sachverhalts behandelt. Die verschiedenen Strategien werden unterschieden nach der Intensität, mit der die Sprecher die Probleme allein oder unter aktiver Einbeziehung des Hörers zu lösen versuchen, sowie nach den Verfahren, mit denen der gesuchte Begriff zu treffen versucht wird. Beide Arbeiten zeigen, daß die Verbalisierung von Wissen vom Hörer mitbeeinflusst wird. Bei Verbalisierungen in einer zweiten Sprache kommt noch hinzu, daß für das vorhandene Wissen die für adäquat gehaltenen Ausdrücke die passenden Worte fehlen, so daß eine Reihe von lexikalischen Strategien angewendet wird. Dennoch ist belegt, daß die geäußerten Worte den gemeinten Sachverhalt nur partiell oder gar nicht treffen. Deshalb sollte eine zweitsprachliche Wortschatzuntersuchung auch solche Aufgaben umfassen, bei denen die Äußerungen und der zugrundeliegende Sachverhalt in die Analyse eingehen können.

Hatch (1978) untersucht Verfahren der Etablierung von Themen zwischen Zweitsprachlernern und Muttersprachlern. Sie zeigt, daß der thematische Rahmen die Verstehensprozesse steuert, indem die perzipierten Äußerungen auf den angenommenen thematischen Rahmen bezogen werden. Nach geglückter Themenetablierung können Äußerungen auf das Wissen bezogen werden, das zu dem Thema aktiviert ist. Durch Verfahren der Zerlegung und interaktiven Absicherung können Sachverhalte behandelt werden, zu deren strikt monologischen Verbalisierung die Zweitsprachler kaum in der Lage wären. Probleme und Mißverständnisse sind bei der Etablierung des Gesprächsgegenstandes und beim Themenwechsel festzustellen.

Rehbein (1982) stellt bei *Worterklärungen* fest, daß das verbalisierbare Wissen von funktionalen Zweckbestimmungen der zugrundeliegenden Sachverhalte strukturiert wird. Auch hier zeigt sich, daß die Verbalisierung, die Tiefe und Breite der Erklärung durch den Gesprächspartner, dessen initiale Anforderung und dessen Aktivitäten während der Verbalisierung beeinflusst wird. Zunächst muß der vom Interviewer genannte Begriff verstanden werden. Bei Äußerungen des Nichtwissens können interaktional angeforderte Hilfen zur erfolgreichen Suche führen. Rehbein zeigt bei einem zweisprachigen türkischen Mädchen, daß es nach der Nennung des deutschen Begriffs "Miete" zunächst das Nichtwissen äußert, dann aber nach Nennung des türkischen Äquivalents "kira" den Begriff auf Deutsch erklären kann. Nach der Rezeption setzt eine innere Suche nach dem genannten Begriff ein. Dies wird oft durch eine Repetition des gesuchten Wortes nach außen gesetzt. Bei der Verbalisierung haben zweisprachige türkische Kinder in Deutschland im Vergleich zu einsprachigen türkischen Kindern in der Türkei und deutschen in Deutschland größere Probleme, die

sprachlich angemessenen Formen für die zweckbezogene Begriffserklärung zu finden. Die letzten beiden Arbeiten unterstreichen, daß ein strikt monolinguales und -logisches Vorgehen bei der Wortschatzuntersuchung der zwei- und zweitsprachigen Alltagspraxis nicht entspricht, da sie die Ressourcen der Probanden auf die produktiven Fertigkeiten in nur einer Sprache beschränkt. Gerade die interaktiv dialogische Bearbeitung erweitert den Handlungsspielraum erheblich. Allerdings gestaltet sich die Analyse zweisprachiger und/oder dialogischer Daten erheblich aufwendiger und komplexer, da systematisch einzubeziehen ist, wie sich der verbalisierte Wortschatz in der Interaktion mit dem Interviewer entwickelt.

Rehbein (1984) untersucht das Verhältnis von Illokution, propositionalem Gehalt und Verbalisierung im alltäglichen Handeln und im institutionalisierten Fremdsprachenunterricht. Der im alltäglichen Kontext gegebene Zusammenhang von Sprechhandlungsziel und der zweckgerichteten Verwendung sprachlicher Mittel in einer Folge von Äußerungen wird im Unterricht aufgelöst. Während im Alltag z. B. im Rahmen einer Kaufhandlung eine Frage geäußert wird, um den Preis einer Ware, an der ein Interesse besteht, in Erfahrung bringen, dient eine solche Frage im Unterricht dazu, dem Lehrer zu zeigen, daß man in der Lage ist, eine grammatisch korrekte Frage zu äußern. Bei Problemen benutzt der Lehrer insbesondere *reparative Handlungsmuster*, um den Schüler zur Äußerung der zielsprachlichen korrekten Form zu veranlassen. Allerdings wird dadurch auch für den Lerner der Zusammenhang zwischen der sprachlichen Form der Äußerungen und ihrer Funktion aufgelöst, so daß in den meisten Fällen kein Lernprozeß stattfindet. Diese Befunde bedeuten für die Elizitierung im Rahmen von Wortschatzanalysen, daß bei zunehmender Distanz zum alltagssprachlichen Zweck - Mittel - Verhältnis auch nicht der im alltäglichen Handeln verfügbare Wortschatz mobilisiert wird, sondern ein dem institutionalisierten Erhebungszweck entsprechender.

Kontext und Mittel des Symbolfeldes:

Aus dem kurzen Forschungsbericht ergibt sich, daß der Zugang des Sprechers zu Elementen des Symbolfeldes zumindest von folgenden Faktoren beeinflußt wird:

- Der *Handlungskontext* steuert die Verfügbarkeit des Zugangs zu Wörtern; so spielt es eine Rolle, ob eine Datenerhebung zum Wortschatz in der Institution Schule oder in der Institution Familie durchgeführt wird. Die Anwesenheit weiterer Personen kann den Zugang ebenfalls beeinflussen: eine Lehrperson ist eben auch immer eine Person institutionalisierter Lehr- Lernprozesse, während Geschwister vertraute Handlungspartner sind, mit denen über den gemeinsamen Lebenszusammenhang auch große Wissensbestände geteilt werden, die auch weitgehend im gemeinsamen Handeln kommunikativ erarbeitet wurden.

- Der *thematische Kontext* aktiviert ein themenbezogenes Wissenspotential, das auf den verschiedenen Strukturtypen des Wissens aufbaut - sowohl partikulärem Erleb-

niswissen wie auch komplexem sprachlichem Musterwissen. Schon bei der Rezeption der vom Interviewer angeforderten Tätigkeit kann so über eine erfolgreiche oder erfolglose Suche nach dem angezielten Wissensbereich entschieden werden.

- Die *Art der angeforderten Verbalisierung* bedingt spezifische Möglichkeiten der Verbalisierung des mobilisierten Wissens. Hier sind sowohl interaktionale Gesichtspunkte bestimmend wie die Möglichkeit, im strikt monolingualen oder bilingualen Modus handeln zu können. Als weiterer Faktor ist die Art der gestellten Aufgabe zu berücksichtigen: das Erzählen selbsterlebter Ereignisse oder die Wiedergabe von gesehenen, bzw. gehörten Geschichten erfolgt in einer Sequenz verketteter Sprechhandlungen, während die Aufforderung, Wörter zu nennen oder etwas zu nennen, was dem Sprecher zu einem Stichwort einfällt, typischerweise zur Produktion unverbundener Wörter oder Satzfragmente führt. Die Elizitierung anhand von Bildvorlagen erfolgt oft mit deiktischen Verweisen auf ein Bildelement und der Nennung eines Symbolfeldausdrucks. Schriftliche Aufgabenstellungen werden mit einem der Schriftsprache entsprechend differenzierten Wortschatz im Unterschied zum eher lockeren Gebrauch umgangssprachlicher Wendungen führt.

Da die verschiedenen Dimensionen des Wortschatzes, über die ein Sprecher zu einem gegebenen Zeitpunkt verfügt, aufgrund der vielfältigen Gebundenheit an die Bedingungen der Handlungskonstellation und der gestellten Aufgabe durch einen Aufgabentyp in nur einem Handlungstyp mit Sicherheit nicht zu elizitieren sind, ist für eine systematische Wortschatzanalyse eine *Mischung* verschiedener *Handlungskonstellationen* und *Aufgabentypen* anzustreben. Angesichts der oben diskutierten Aspekte und Erfahrungen aus anderen Wortschatzanalysen sollte eine systematische Erhebung folgende Aufgabentypen umfassen:

- freie Assoziationen ohne Bezug im aktuellen Handlungskontext;
- Verbalisierung einer Bildvorlage (Einzelbilder, Geschichte);
- Verbalisierung irgendeiner einer bekannten 'Geschichte';
- Verbalisierung irgendeiner erinnerten einzelnen Begebenheit;
- Verbalisierung einer selbsterlebten (regelmäßigen) Handlung (Spielen, Einkaufen).

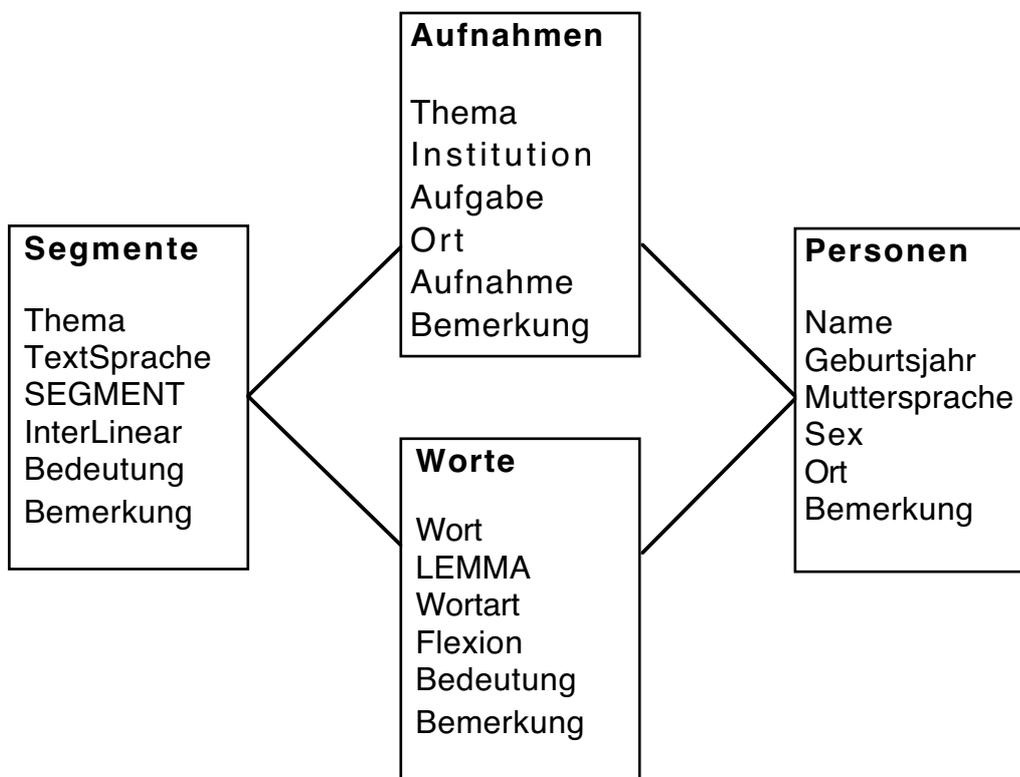
Bei mehrsprachigen Wortschatzanalysen sollten muttersprachliche Sprecher die Interviews in der jeweils erhobenen Sprache durchführen. Da die Aufgaben aus Gründen der Vergleichbarkeit im Normalfall parallel in beiden Sprachen sind, wäre bei nur einem Interviewpartner dessen Zweitsprache, bzw. weniger gut beherrschte Sprache, zurückgedrängt, da der Proband dem Interviewer etwas mitteilen soll, was ihm in der anderen Sprache so oder ähnlich, bzw. detaillierter oder lebendiger schon mitgeteilt wurde. Der Bekanntheitseffekt betrifft auch die von Rehbein (1984) angesprochene Relation von Sprechhandlungszielen: während für die erste Aufgabenbewältigung wahrscheinlich ein gewisser authentischer Mitteilungsgehalt angenommen werden kann, entfällt dieser Handlungszweck bei der Wiederholung in der zweiten Sprache. Diese Verbalisierung erfolgt somit auf einer anderen Handlungs-

basis und ist deshalb nur noch indirekt über vermittelnde Interpretationsschritte mit den anderen Daten vergleichbar.

Anforderungen an eine kontextsensitive Datenerfassung:

Die Berücksichtigung des Kontextes bei der Wortschatzanalyse erfordert die Erfassung und Speicherung von Daten folgender Kontextbereiche: (1) des sozial-biographischen Kontexts, (2) des äußerungsexternen Erhebungskontexts (Verwendungszusammenhang der sprachlichen Mittel, Aufgabentyp, Äußerungsmodus) und (3) des äußerungsinternen Verwendungskontexts (isoliert - frei - kombiniert).

Da sich die kontextualisierte Wortschatzanalyse im Kern auf die Mittel des Symbolfeldes bezieht, d. h. auf 'Wörter', bilden sie die kleinste Analyse- und Beschreibungseinheit. Der Zugriff auf die Kontextdaten wird dadurch hergestellt, daß die Wörter jeweils mit den Sätzen, bzw. Äußerungssegmenten verbunden werden, in denen sie enthalten sind, d. h. sie werden in die höhere Segmenteinheit eingebettet. Gleichzeitig sind die Segmente einerseits mit Angaben zu den Personen und ihren biographischen Daten verbunden und andererseits mit Angaben zu den Erhebungs-, bzw. Verwendungskontexten. Im Kern ergibt sich ein Verbund von folgenden vier miteinander verbundenen Dateien: (1) Worte, (2) Segmente, (3) Personen und (4) Aufnahmen. Schematisch lassen sich die Beziehungen wie in Figur 1 darstellen:



Figur 1: Schematische Darstellung der Beziehungen zwischen den Dateien

§6 Aufbereitung der Daten, Lemmatisierung

Häufigstes analytisches Verfahren bei der Wortschatzuntersuchung sind globale Relationen des Type / Token Verhältnisses (s. z. B. die Makro Perspektive im ESF-Projekt (Broeder & Extra & van Houten (1989))). Bei der computergestützten Analyse werden für diese Arbeiten häufig Konkordanzprogramme benutzt⁴, die automatisch den 'Text', d. h. die Äußerungen eines Probanden in Wortlisten zerlegen, diese Wortlisten alphabetisch sortieren und gleiche Wortformen auszählen. Vor dem letzten Analyseschritt werden die Wortlisten in Worttypenlisten überführt, 'Lemmatisiert', wobei flektierte Vollformen auf Stammformen zurückgeführt und je nach Untersuchungszweck die Wortklassenzugehörigkeit bestimmt und notiert wird (Drewek (1989: 10) gibt ein Beispiel eines computergestützt erstellten lexikalischen Frequenzprofils). Beispiel 1 ist ein Auszug aus einer schriftlichen Textproduktion einer türkischen Schülerin zum Thema "Wochenende" auf deutsch und auf türkisch; Beispiel 2 illustriert eine daraus abgeleitete Wortliste mit Belegen in der Reihenfolge ihres Auftretens, den zugeordneten Lemmata und den zugewiesenen Wortklassen⁵.

(B 1) Text - Auszug: Suzan Aygül

deutsch:

- 0) Wochenende
 1) samstag morgen waren wir 9.30 in Lübek in Markauf
 2) und hatten etwas gekauft.

türkisch:

- 0) Hafdasonu
 1) Cumartesi günü sabahlayın 9.30 da Lübekedeki Markaflar gidikk

(B 2) Lemmatisierte Worttypenliste - Auszug: Suzan Aygül

deutsch:			türkisch:		
Beleg	Lemma	Desk.	Beleg	Lemma	Desk.
Wochenende	WOCHENENDE	NOM	Hafdasonu	HAFTA	NOM
samstag	SAMSTAG	NOM	Hafdasonu	SON	NOM
morgen	MORGEN	ADV	Cumartesi	CUMARTESI	NOM
waren	SEIN	VER	günü	GÜN	NOM
wir	UNS	PRO	sabahlayın	SABAHLEYİN	ADV
9.30			9.30		

⁴ Als Standard für Personalcomputer mit dem Betriebssystem MS-DOS kann Micro-OCP angesehen werden.

⁵ Aus Gründen der Übersichtlichkeit wurden die deutschen und türkischen Belege in zwei parallelen Spalten angeordnet.

in	IN	PRP	da	DA	ADV
Lübek	LÜBEK	NOM	Lübekedeki	LÜBEK	NOM
in	IN	PRP	Markaufe	MARKAUF	NOM
Markauf	MARKAUF	NOM	gidikk	GITMEK	VER
und	UND	KON	ve	VE	KON
haten	HABEN	VER	ordan	ORADAN	ADV
etwas	ETWAS	PRO	bişeyler	BIR	NUM
gekauft.	KAUFEN	VER	bişeyler	ŞEY	NOM
			aldık	ALMAK	VER

Die Auszählung erfolgt dann anhand einer Liste, die nur noch die Lemmata und die Wortklassendescriptoren enthält. Der Bezug zwischen einem bestimmten, mehrfach enthaltenen Lemma und den einzelnen Belegen ist nicht mehr gegeben. Es kann auf dieser Bearbeitungsstufe z. B. nicht mehr (ohne aufwendige manuelle Textvergleiche) rekonstruiert werden, in welchen Varianten ein Lemma im zugrundeliegenden Text belegt ist. Dadurch können z. B. keine Aussagen darüber gemacht werden, ob z. B. im Anschluß an Intervieweräußerungen bestimmte Formen standardnäher geäußert werden als ohne. Dies könnte ein Anhaltspunkt für weitere Untersuchungen zur Speicherung der Wortformen im mentalen Lexikon des Lerners sein. Derart interaktionale orientierte Analysen erfordern bei diesem Vorgehen getrennte zusätzliche Arbeitsschritte.

Ein weiteres Problem des dekontextualisierenden Vorgehens (zu diesem Terminus s. Grißhaber 1988) ist darin zu sehen, daß die aus ihrem Äußerungskontext herausgelösten Wörter nicht mehr im ursprünglichen Äußerungszusammenhang betrachtet werden können. Die Analyse arbeitet dann ausschließlich mit analytisch aufbereiteten normierten Einheiten, deren spätere Überprüfung nahezu unmöglich ist, wenn die Zuweisung einmal durchgeführt worden ist.

Die kurzen, in der Schreibweise z. T. stark variierenden Beispiele belegen weiterhin, daß die Aufbereitung der Daten für die Analyse in jedem Fall eine bewußte Bearbeitung erfordert. Die häufig nicht standardgemäßen Formen schließen beim derzeitigen Stand von Parsern eine automatische Wortformenanalyse und Lemmatisierung aus, da sie aufgrund der häufigen normabweichenden Realisierungen derzeit kaum ohne extensives Monitoren durchzuführen wären.

Die Berücksichtigung des Kontextes bei der Wortschatzanalyse bedingt die Beachtung folgender Grundsätze bei der Lemmatisierung der Äußerungen:

- **Speicherung der Äußerungen aller beteiligten Personen:**

Alle Äußerungen werden gespeichert, nicht nur die der Probanden. Nur unter Einbeziehung der Interaktionsstruktur und aller während der Erhebung erfaßten Daten erlaubt bei der späteren Analyse die Integration kommunikativer Prozesse der Wort-

findung und Verbalisierung.

- ***Speicherung ganzer Äußerungen und nicht nur einzelner Wörter:***

Die Äußerungen sollten jeweils als ganze gespeichert werden und nicht nur als sequentielle Liste von Wörtern. Mit relationalen Datenbanksystemen ist es möglich, die Daten in Form von sequentiell angeordneten Listen zu speichern, deren Wörter jeweils mit einem Datensatz der betreffenden Wörter verbunden sind. Dadurch ist es möglich, z. B. bei der Suche nach Äußerungen, die ein bestimmtes Merkmal enthalten, jeweils direkt, d. h. ohne Umweg, die gesamte Äußerung zu berücksichtigen, in die die betreffenden Wörter eingebettet sind.

- ***Speicherung von Wortformen zusammen mit den Lemmata:***

Die computergestützte kontextualisierte Analyse erfordert den Zugriff sowohl auf die von den Probanden benutzten einzelnen Wortformen wie auf die analytisch bestimmten lemmatisierten Grundformen.

- ***Möglichst geringe Interpretation der Daten bei der Deskription:***

Dem Ziel einer kontextualisierenden Datenanalyse steht das übliche Vorgehen entgegen, die Daten schon möglichst früh mit klassifizierenden Deskriptoren zu versehen und die Analyse vorwiegend auf die Deskriptoren, statt auf die Daten zu stützen. Dagegen wird angestrebt, die Daten nach Möglichkeit so zu erfassen, daß die Analyseprozeduren direkt auf die unveränderten Daten anwendbar sind, klassifizierende Deskriptionen dagegen zu reduzieren. So weit eine klassifizierende Deskription erfolgt, sollte sie sich so weit wie möglich, auf äußere, schnell entscheidbare Merkmale beziehen.

§7 Modular getrennte Deskription

Deskriptive Informationen lassen sich nach verschiedenen Verfahren kodieren und speichern. Zu berücksichtigende Faktoren sind vor allem: klare Unterscheidung zwischen den Deskriptoren, Speicherbedarf bei 'sprechenden', ausgeschriebenen Formen, mnemotechnische Prägnanz für die Datenbeschreibung, Lesbarkeit bei der Auswertung der Datensätze. Diese gegensätzlichen Ansprüche lassen sich auch bei sorgfältiger Konstruktion nicht zur Deckung bringen. Dominanter Faktor war längere Zeit hindurch aufgrund der hardwareseitig gegebenen Restriktionen der Speicherbedarf, so daß vor allem zur Platzersparnis kompakte Kodiersysteme benutzt wurden.

Etliche Konventionen versuchen, nicht zuletzt aus Gründen der Platzersparnis, möglichst viele Informationen in kompakten Deskriptoren zu komprimieren.

Ein Beispiel dafür ist die Kodierung von Satzmustern im dtv Wörterbuch (1978: 30f.). Die dreistelligen Ziffern zwischen 300 und 580 repräsentieren verbabhängige Satzmuster. Während die erste Ziffer jeweils für eine bestimmte Klasse von Mustern steht, die 5 z. B. für Satzmuster mit obligatorischem

Akkusativobjekt, repräsentieren die übrigen zwei Ziffern kein derart modular voneinander unabhängiges Muster. Eine 1 an letzter Stelle kann so für eine Ortsbezeichnung stehen, kann aber auch für andere Konstituenten, wie ein nicht lokales Präpositionalobjekt stehen, z. B. in 801 für "es wimmelt von Ameisen".

Ein solches System ist gerade aufgrund seiner Kompaktheit für die vorliegenden Zwecke wenig geeignet. Probleme ergeben sich vor allem in folgenden Bereichen: bei der Suche nach unberücksichtigten Merkmalskombinationen und bei der Änderung, bzw. Erweiterung der Deskriptoren. Wenn nach Merkmalskombinationen gesucht werden soll, die bei der Kodierung nicht vorgesehen worden waren, wird die Formulierung geeigneter Abfrageprozeduren schwierig. Sollen z. B. nach dem Muster des dtv Wörterbuchs nach lokalen Ortsbestimmungen gesucht werden, sind aufwendige und komplexe Abfragen notwendig, da alle Kodiervarianten zu berücksichtigen sind, die das gesuchte Merkmal bezeichnen. Der explorative Charakter der KWA bringt es mit sich, daß nicht ein einmal festgelegtes Set von Deskriptoren wie ein Raster über die Daten gelegt werden kann. In Abhängigkeit vom jeweiligen Untersuchungszweck und der konkreten Fragestellung ergeben sich unterschiedliche Beschreibungsschwerpunkte. Deshalb muß das gewählte Kodiersystem mögliche Änderungen und/oder Erweiterungen der Deskriptoren unterstützen. Das läßt sich am leichtesten und sichersten durch ein modular aufgebautes Design bewerkstelligen, bei dem jede Merkmalsausprägung durch einen eigenen Deskriptor erfaßt wird, der nur Angaben zu dem betreffenden Merkmal enthält.

Ausgangsbasis der wortbasierten Deskription sind folgende Daten und Merkmale:

- (0) WORT und LEMMA
- (1) Angaben zur Wortart (s. u.)
- (2) Angaben zur Wortbildung (z. B. Komposition, Derivation)
- (3) Angaben zu unveränderlichen Merkmalen (z. B. Genus im Deutschen)
- (4) morphologische Angaben zu Suffixen (z. B. Numerus, Kasus, Tempus)
- (5) Angaben zum Worttyp
- (6) Angaben zur Wortsemantik
- (7) Angaben zu wortübergreifenden Zusammenhängen
- (8) Verbindung zu Kontextdaten, d. h.:
 - zu den segmentierten Äußerungen, in denen die Worte enthalten sind
 - zu den Sprechern mit biographischen Daten
 - zu den Erhebungskontexten (z. B. Aufgabe, Institution)

Das Hauptarbeitsfenster und die Gruppierung der Daten

Bei der technischen Realisierung der Datenerfassung und -beschreibung mittels einer relationalen Datenbank werden diese Daten in vier selbständigen miteinander verknüpften Dateien gespeichert (s. o.). Das Fenster für die Dateneingabe basiert auf der Datei Segmente, in der die Äußerungen auf syntaktischer Ebene präsentiert werden.

Die Informationen sind in drei Blöcken gruppiert:

- (1) Kopfteil mit Kontextdaten zur Aufnahmekonstellation und Aufgabenstellung;
- (2) Segmente;
- (3) den wortbezogenen Wörtern/Lemmata mit Deskriptoren.

①

sid#	2	seg#	1	fol#		WENDE	Sgl	Met	SSpr	D.s		TSpr	D.s		Met	OS	TR
------	---	------	---	------	--	-------	-----	-----	------	-----	--	------	-----	--	-----	----	----

②

SEGM	in den alkauf sint vier																
IntLin																	
SBem																	

③

w#	wid#	LEMMA	WORT	wart	wsem	wbld	wkas	wnum	wgen	vtmp	wtyp	KST
1	1		in									

(1) Kopfteil:

Hier sind die für die Auswertung der Daten am häufigsten benötigten Daten zusammengefaßt. Sie umfassen Daten zum erfaßten Segment sowie zum äußerungsexternen Kontext. Aufgrund der relationalen Verknüpfung der Datensätze sind alle erfaßten Daten jederzeit verfügbar. Die Angaben im Haupteingabefenster umfassen:

sid#	2	seg#	1	fol#		WENDE	Sgl	Met	SSpr	D.s		TSpr	D.s		Met	OS	TR
------	---	------	---	------	--	-------	-----	-----	------	-----	--	------	-----	--	-----	----	----

Angaben zum Segment:

- durchlaufende Indexnummer über alle Segmente (Feld **sid#**, im Beispiel 2)
- laufende Indexnummer im aktuellen Kontext (Feld **seg#**, im Beispiel 1)
- segmentinterne Bezugsnummer (z. B. Bilder) (Feld **fol#**, im Beispiel -)

Angaben zu äußerungsexternen Kontextdaten:

- Deskriptor für Aufnahmekontext (Feld **Thema**, im Beispiel: WENDE)

Angaben zum sozialbiographischen Kontext:

- Personenangaben mit Sigle (Feld **Sgl**, im Beispiel: Met)
- Muttersprache (Feld **MSpr**, im Beispiel: TR)
- Geschlecht (Feld **Sex**, im Beispiel: m)
- Gruppe (Feld **Grp**, im Beispiel: OS)

(2) Die Segmente:

Dieser Teil enthält die nach syntaktischen Kriterien segmentierten Äußerungen:

SEGM	in den alkauf sint vier																
IntLin																	
SBem																	

- das zu beschreibende Segment (Feld **SEGM**, im Beispiel: in den alkauf sint vier)
- eine Interlinearübersetzung (Feld **IntLin**, im Beispiel -)
- ein Feld für Bemerkungen (Feld **SBem**, im Beispiel -)

(3) Wortbezogene Daten und Deskriptoren:

Die matrizenförmig angelegten Felder sind aus der Wortdatei und sind mit dem jeweiligen Segment-Datensatz verbunden. Die Daten können in dieser Ansicht betrachtet und bearbeitet werden. Sie enthalten neben fortlaufenden Indexnummern den Kern der wortbezogenen Daten und Deskriptoren:

w#	wid#	LEMMA	WORT	wart	wsem	wbld	wkas	wnum	wgen	vtmp	wtyp	KST
1	1	in	in	PRP	lok							
2	2	der	den	ARD								
3	3	Altkauf	alkauf	NOM	lok						e	
4	4	sein	sint	VER				•				
5	5	wir	vier	DXS				•				

- durchlaufende Indexnummer über alle Wörter (Feld **w#**, im Beispiel 1)
- laufende Indexnummer im aktuellen Segment (Feld **wid#**, im Beispiel 1)
- ikonische Darstellung der Wortart (Feld **wartBild**, im Beispiel >)
- das zugewiesene Lemma (Feld **LEMMA**, im Beispiel in)
- das zu beschreibende Wort (Feld **WORT**, im Beispiel in)
- der Deskriptor für die Wortart (im Feld **wart**, im Beispiel PRP)
- der Deskriptor für die Wortsemantik (im Feld **wsem**, im Beispiel lok)
- der Deskriptor für die Wortbildung (im Feld **wbld**, im Beispiel -)
- der Deskriptor für den Kasus (im Feld **wkas**, im Beispiel -)
- der Deskriptor für den Numerus (im Feld **wnum**, im Beispiel -)
- der Deskriptor für das Genus (im Feld **wgen**, im Beispiel -)
- der Deskriptor für die Tempusmorphologie (im Feld **vtmp**, im Beispiel -)
- der Deskriptor für Angaben zur Wortstruktur (im Feld **wtyp**, im Beispiel -)
- der Deskriptor für wortübergreifende Angaben (im Feld **KST**, im Beispiel -)

§8 Wortbezogene Deskriptoren

Entsprechend der zugrundeliegten funktional-pragmatischen Sprachtheorie konzentriert sich die Beschreibung und Analyse auf folgende detailliert beschriebene Wortarten: Adjektiv, Adverb, Substantive und Verben. Die übrigen Wörter werden überwiegend lediglich ihrer Wortart nach bestimmt und je nach Untersuchungszweck um weitere Angaben ergänzt.

Die Beschreibung erfolgt auf Wortebene. Jeder selbständige Ausdruck wird beschrieben, auch wenn ein Wort in zwei getrennten Ausdrücken realisiert wird. Ein Beispiel für diskontinuierlich realisierte Ausdrücke sind z. B. trennbare Verben im Deutschen, bei denen Stamm und Präfix zusammen oder getrennt auftreten können (auf| stehen, steht ... auf). Der ebenfalls mögliche Zugang auf Morphemebene, der sich auf den endungslosen Stamm plus aktuell verbundene Morpheme bezieht, wird nicht realisiert, obwohl es eine sehr differenzierten Beschreibung gewährleistet, die gerade bei

der agglutinierenden Struktur des Türkischen die Transparenz für vergleichende Studien erhöhen würde. Die äußerst aufwendigen Deskriptionsprozeduren stehen einer generellen Anwendung dieses Systems für die Zwecke der vorliegenden Untersuchung entgegen. Ein Wort wie "aufgestanden" würde in die drei Morpheme "steh", "auf" und "ge - en +Ablaut" zerlegt und getrennt beschrieben. Die dabei erfaßten Informationen über den morphologischen Aufbau sind für die Wortschatzanalyse nicht erforderlich. In Einzelfällen ist das Vorgehen durchaus beschreibungsadäquat, z. B. für den Bereich komplex gebildeter Mittel der operativen Prozedur (z. B. zusammengesetzte Verweiswörter wie "da | her").

Die parallele Beschreibung von Äußerungen in zwei so verschiedenen Sprachen wie Deutsch und Türkisch ist schon beim Design des Systems zu berücksichtigen. So ist für die praktische Arbeit mit den Daten anzustreben, daß in beiden Sprachen weitgehend parallele Suchoperationen durchgeführt werden können. Dazu werden auf der Basis von funktionalen Bestimmungen ähnliche Gruppen mit gleichen Deskriptoren erfaßt, auch wenn im streng strukturalistischen Sinn aufgrund unterschiedlicher Rahmenbedingungen kein tertium comparationis existiert. Einige Wortarten gibt es nur in einer der beiden Sprachen, z. B. Anapher oder Artikel im Deutschen.⁶

Andererseits repräsentiert mit Blick auf die computergestützte Erfassung und Auswertung nicht jeder Deskriptor eine eigene Wortart. So sind z. B. die Artikel aus praktischen Gründen schon auf der Deskriptorebene in definite und indefinite Artikel unterteilt (ARD - ARI), obwohl beide derselben Wortart angehören. Die beiden identischen Startzeichen (AR) erlaubt ihre Zusammenfassung in einer einheitlichen Gruppe bei gleichzeitiger Möglichkeit zur differenzierten Auswertung aufgrund der unterschiedlichen Endzeichen. Dagegen werden die üblicherweise als Pronomina zusammengefaßten und gemeinsam behandelten Anaphern und Deiktika aufgrund ihrer unterschiedlichen funktionalen Leistung als verschiedene Wortarten erfaßt und mit verschiedenen Deskriptoren beschrieben. Innerhalb der deiktischen Mittel wird erfolgt dann wie bei den Artikeln über differente Endzeichen eine Untergruppierung nach einfachen und zusammengesetzten Deiktika (DXS - DXZ).

⁶ Dies bedeutet natürlich nicht, daß die Funktionen, die diese Mittel erfüllen, im Türkischen nicht ausgedrückt würden.

Deskriptoren (nach Wortarten):

Desk.	Wortart	Deutsch	Türkisch	Prozedur
ADJ	● Adjektiv	<i>schön</i>	güzel	NN
ADV	● Adverb	<i>schnell</i>	h/z l/	NN
ANA	Anapher	<i>sie</i>	–	OP
ARD	Artikel, definit	<i>die</i>	–	OP
ARI	Artikel, indefinit	<i>ein</i>	–	OP
AUG	Augment	<i>ne?</i>	de©il mi?	EP
DXS	Deixis, simplex	<i>das, jetzt, du</i>	bu, ,imdi, sen	DX
DXZ	Deixis, zusammengesetzt	<i>danach</i>	ondan sonra	DX
IDF	Indefinitpr.	<i>jeder</i>	her	OP
INT	Interjektion	<i>ohh</i>	ooo	EP
KOK	Konjunktion, koordinierend	<i>und</i>	ve	OP
KOS	Konjunktion, subordinierend	<i>als</i>	[Suffixe]	OP
NOM	● Substantiv	<i>Kind</i>	çocuk	NN
NUM	● Numeral	<i>zwei</i>	i k i	NN
PAR	Partikel	<i>ja</i>	da	OP
POS	Possessivum	<i>mein</i>	[Suffix], benim	OP
PRP	Präposition, Postposition	<i>auf, für</i>	[Suffix], için	OP
QUE	Fragewort, simplex	<i>was?</i>	ne?	OP
QUZ	Fragewort, zusammengesetzt	<i>woher?</i>	–	OP
REL	Relativum	<i>deren</i>	–	OP
VER	● Verb	<i>kommen</i>	gelmek	NN
VEA	Verb, auxiliar	<i>haben sein werden</i>	–	OP
VEM	● Verb, modal	<i>wollen</i>	–	NN
VEP	● Verb, Partizip	<i>gekommen</i>	[Suffixe: -(y)An, ...]	NN
VEG	● Verb, Gerundium	–	[Suffixe: -ken, (y)ArAk, ...]	NN
VZS	Verbzusatz	<i>(kommt) her</i>	–	–

Legende: NN nennende Prozedur
 DX deiktische Prozedur
 ● Kern der Symbolfeldausdrücke
 OP operative Prozedur
 EP expeditive Prozedur

§9 Einzelfragen der Lemmatisierung

Im Mittelpunkt der zweisprachigen Wortschatzanalyse steht die lexikalische Bedeutung der Elemente in ihrem Kontext. Die Deskription zielt auf die Erfassung der Lemmata. Dagegen treten syntaktische Verhältnisse und ihre durch Suffixe realisierten Beziehungen zurück. Im Deskriptionsprozeß wird den flektierten Vollformen die unflektierte Grundform zugewiesen, bei Verben z. B. der rekonstruierte Infinitiv. Diese interpretativ erschlossene Grundform wird zusätzlich zu den von den Spre-

chern verwendeten Vollformen als Lemma eingetragen.

Allgemeine Deskriptionsgrundsätze

Grundsätzlich wird nach dem Prinzip verfahren, auf den Gebrauch expliziter Deskriptoren nach Möglichkeit zu verzichten. Dadurch bleiben die Primärdaten so weit wie möglich erhalten und mit der Reduktion der expliziten Merkmalszuschreibung wird auch die Gefahr reduziert, eine fehlerhafte Deskription vorzunehmen. Nach diesem Prinzip wird insbesondere bei der Bestimmung morphologischer Merkmale verfahren. Obwohl die Matrix eine vollständige morphologische Beschreibung der Wörter erlaubt, wird nach Möglichkeit darauf verzichtet und gegebene 'Default'-Werte genutzt.

Kasus:

Normalerweise wird der Kasus nicht eingetragen. Zum einen ist im Deutschen ein Kasus oft nicht durch ein Suffix an der sprachlichen Oberfläche wahrnehmbar, so daß in vielen Fällen von der vorliegenden Form eines Wortes nicht eindeutig auf den Kasus geschlossen werden kann. Zum anderen sind Kasusmarkierungen kein zentraler Bestandteil der kontextualisierten Wortschatzanalyse. In den Fällen jedoch, in denen der Kasus eine funktionale Rolle auch auf der Wortschatzebene spielt, wird er vermerkt, z. B. bei der bestimmten "izafet"-Verbindung.

Numerus:

Worte im Singular werden im Feld für Numerus-Werte nicht markiert. Lediglich Worte im Plural werden markiert. Da in diesem Feld nur die dichotomischen Werte "S" oder "P" möglich sind, wird das auffällige "•" Zeichen für die Pluralbeschreibung im Feld W-Num verwendet. Es ist in der tabellarischen Darstellung leichter perzipierbar als ein alphanumerischer Deskriptor und erlaubt dadurch eine größere Übersichtlichkeit.

Behandlung von inkorrekten Ausdrücken:

Auch im Hinblick auf inkorrekte Ausdrücke, vor allem Genusfehler im Deutschen, wird das oben formulierte Prinzip der geringstmöglichen Deskription verfolgt. Deshalb werden generell nur solche Inkorrektheiten vermerkt, die direkt an der sprachlichen Oberfläche erkennbar sind. In diesem Fall wird ein Schrägstrich in das Feld eingetragen, in das das Beschreibungsmerkmal eingetragen wird. Für falschen Genusgebrauch wird also ein Schrägstrich eingegeben. Der korrekte Wert wird direkt hinter dem Schrägstrich eingegeben. Der inkorrekte, falls er eindeutig bestimmbar ist, was oft unmöglich ist, wird direkt vor dem schrägstrich eingegeben. Ein inkorrekt Genusgebrauch bei den Worten (2) "den" und (3) "bet" wird demnach mit folgenden Einträgen erfaßt:

w#	wid#	LEMMA	WORT	wart	wsem	wbld	wkas	wnum	wgen	vtmp	wtyp	KST
6	1	in	in	PRP	lok							
7	2	das	den	ARD				m/n				
8	3	bet	bet	NOM				m/n				
9	4	gehen	gegangen	VEP	bew							

Lemmatisierung von Verben

Die Verben werden in folgende drei Gruppen eingeteilt:

- (a) Vollverben (VER) und Auxiliare (VEA), bzw. Modalverben (VEM);
- (b) Infinite und finite Formen (VER), Partizipien (VEP), Gerundien (VEG);
- (c) getrennt vorkommenden Verbpräfixen (VZS).

Diese Gruppierung berücksichtigt die Mittelstellung der Partizipien im Deutschen. Für das Türkische ist die Unterscheidung notwendig, da Partizipien und Gerundien gleichzeitig komplexe syntaktische Funktionen wahrnehmen, die im Deutschen mit subordinierten Nebensätzen wiedergegeben werden können (s. Rehbein 1989). Der Aufbau der Deskriptoren erlaubt auch hier - wie bei den Artikeln - die Bildung von Gruppen. Durch die beiden Startzeichen (VE) sind Verben als Elemente einer Gruppe erfaßbar. Auf der Grundlage der unterschiedlichen Endzeichen können mit algebraisch gebildeten Suchausdrücken differenzierte Auswertungen durchgeführt werden, bei denen einzelne Untergruppen ein- oder ausgeschlossen sind.

Inkorporierte Suffixe im Türkischen

Für das Türkische sind inkorporierte Suffixe typisch, z. B. Gerundien mit dem Suffix -ince. In diesen Fällen wird dem Verb nach dem oben beschriebenen Vorgehen der Infinitiv als Lemma und der entsprechende Deskriptor zugeordnet. Bei einem Gerundium wird zusätzlich zum Deskriptor VEG der Suffix in das Feld "W-Typ" eingetragen (im Bsp. (1) "-ince"):

#	Lemma	Wort	W-Art	W-Sem	W-Bld	W-Typ	KST
1	girmek	girince	VEG	bew	-	-ince	-
2	bilmek	biliyormusun	VER	sen	-	-	QUE

Ähnlich wird das inkorporierte Fragesuffix -mI- bei finiten Verben behandelt, das je nach vorausgehendem Vokal in den Formen -mi-, -m/-, -mu- und -mü- auftritt. Deskriptionsgrundlage ist in diesem Fall der Deskriptor VER für finite Verben. Zusätzlich zu dieser Bestimmung muß das Fragesuffix erfaßt werden, dem im Deutschen meist ein eigenes Fragewort entspricht. Zur Vergleichbarkeit mit deutschen Daten wird es nicht als Suffix eingetragen, sondern mit dem Deskriptor "QUE" für

Fragewörter im Feld KST erfaßt (s. Bsp. (2)).

Behandlung komplexer und diskontinuierlicher Ausdrücke

Diskontinuierlich realisierte Wörter, z. B. finite zusammengesetzte Verben mit trennbarem Präfix (z. B. "stand ... auf", (Nr. 3, 4)) werden nach dem Prinzip der wortbezogenen Lemmatisierung auf der Wortebene als zwei Wörter unter einem Lemma erfaßt.

#	Lemma	Wort	W-Art	W-Sem	W-Bld	W-Typ
1	stehen	stand	VER	pos	–	–
2	aufstehen	aufstehen	VER	–	–	–
3	aufstehen	stand	VER	–	–	–
4	-	auf	VZS	–	–	–
5	sein	ist	VEA	–	–	–
6	aufstehen	aufgestanden	VEP	–	–	–

Auf der Ebene der Lemmata wird für das Beispiel "stand ... auf" im Feld Lemma der Infinitiv "aufstehen" mit Präfix eingetragen (s. (3)). Zusammen mit diesem Lemma wird im Feld W-Art der Deskriptor VER eingetragen. Die diskontinuierliche Realisierung von Stamm und Präfix wird direkt im Lemma durch einen senkrechten Strich zwischen Präfix und Stamm repräsentiert. Der Präfix "auf" (Nr. 4) wird als eigenes Wort ohne Lemmaeintrag mit dem Deskriptor VZS für Verbpräfixe gekennzeichnet. Durch dieses Vorgehen ist sichergestellt, daß Realisierungen mit Präfix, z. B. "aufgestanden" (s. (6)), dasselbe Lemma erhalten. Letzteres erhält den Deskriptor VEP für Partizipien. Durch den senkrechten Strich im Lemma können alle Wörter mit trennbarem Präfix gesucht werden und es können mittels formaler Operationen differenziert Analysen nach Präfixen oder Stämmen durchgeführt werden.

Wortübergreifende Ausdrücke: türkische izafet-Bildungen

Eine Deskriptionsbesonderheit in beiden Sprachen stellt die Behandlung von wortübergreifenden Ausdrücken dar. Wortgruppen sind zwei oder mehr zusammengehörende Wörter, als idiomatischer Ausdruck oder als Redewendung. In diesem Fall ist die allgemeine Bindungsfähigkeit und Bedeutung der einzelnen Wörter eingeschränkt oder aufgehoben. Nach den oben schon vorgestellten Prinzipien werden diese Ausdrücke einmal auf der Ebene der Vollformen in eigenständigen Einträgen erfaßt, andererseits werden die einzelnen Wörter der Wortgruppe so beschrieben, daß ihre Zusammengehörigkeit erhalten bleibt.

Ein Beispiel für derartige Wortzusammensetzungen im Türkischen ist z. B. "hafta sonu" (Wochenende; wörtlich: Woche Ende-ih). Es ist aus den zwei Substantiven "hafta" (Woche) und "son" (Ende) gebildet. Das zweite Wort "son" trägt das Posses-

sivsuffix "I" in der durch die Vokalharmonie bedingten Form "u". Diese Wortgruppe, im Türkischen "izafet"-Verbindung⁷ genannt, wird analog zu diskontinuierlich realisierten Präfixverben mit einem Lemma, aber zwei Worteinträgen erfaßt. Deutsche Komposita werden dagegen wie die kontinuierlich gebrauchten Präfixverben nur mit einem Wort- und Lemmaeintrag erfaßt. In beiden Fällen wird im Feld W-Bld für Wortbildungseinträge der Bildungstyp ("k": Komposition, Zusammenstellung) und die Art der Grundwörter ("NN": Substantiv Substantiv) vermerkt:

#	Lemma	Wort	W-Art	W-Bld	W-Kas	W-Typ	KST
1	hafta sonu	hafta	NOM	k.NN	–	–	izf
2	–	sonu	NOM	–	–	POS	izf
3	Wochenende	Wochenende	NOM	k.NN	–	–	–
4	hafta	haftanın	NOM	–	2	–	izf
5	son	sonu	NOM	–	–	POS	izf

Eine weitere "izafet"-Verbindung⁸ des Türkischen ist in (4)-(5) mit dem Ausdruck "haftanın sonu" (das Ende der Woche; wörtlich: Woche-Genitiv Ende-ihr) dargestellt. Im Unterschied zur unbestimmten Wortverbindung ist das erste Substantiv im Genitiv, erfaßt durch die Ziffer 2 im Feld W-Kas. Durch die unterschiedlichen Einträge in diesem Feld lassen sich die verschiedenen "izafet"-Typen bei der Auswertung auseinanderhalten. Der entsprechende deutsche Ausdruck "das Ende der Woche" wird in vier getrennten Wort- und Lemmaeinträgen erfaßt.

Wortübergreifende Ausdrücke: zusammengesetzte Verweis- und Fragewörter

Bei Erzählungen werden häufig zusammengesetzte Verweiswörter wie "o zaman" (zu jener Zeit, dann) verwendet. "O zaman" ist aus dem Deiktikum "o" (da, (1)) und dem Substantiv "zaman" (Zeit, (2)) gebildet und wird analog zur Erfassung von Wortgruppen mit einem Lemma in zwei Wortfeldern erfaßt:

#	Lemma	Wort	W-Art	W-Bld	W-Kas	W-Typ	KST
1	o zaman	o	DXS	k.DN	–	–	DXZ
2	–	zaman	NOM	–	–	–	DXZ
3	ondan sonra	ondan	DXS	k.DP	5	–	DXZ
4	–	sonra	PRP	–	–	–	DXZ
5	bittikten sonra	bittikten	VEP	k.VP	5	-dik	DXZ
6	–	sonra	PRP	–	–	–	DXZ
7	danach	danach	DXZ	–	–	–	–

⁷ "Belirtisiz tamlama" genannt, d. h. "unbestimmte Verbindung".

⁸ "Belirtili tamlama" genannt, d. h. "bestimmte Verbindung".

Ähnlich werden auch zusammengesetzte Verweiswörter wie "ondan sonra" (danach) erfaßt. Es ist aus dem Deiktikum "o" im Ablativ (da, (3)) und der Postposition "sonra" (nach, (4)) gebildet. Deshalb wird bei ondan zusätzlich im Feld für Kasuseintragungen der Kasus (u) vermerkt. Analog dazu wird der deverbale Ausdruck "bittikten sonra" erfaßt, dessen erster Teil aus dem Verb "bitmek" (zu Ende gehen) mit dem Partizipsuffix "-dik" im Ablativ (5) und der Postposition "sonra" (nach, (6)) gebildet ist. Wie bei den oben vorgestellten Gerundien wird auch das Partizipsuffix "-dik" im Feld *W-Typ* eingetragen.

Im Deutschen werden entsprechende, zusammengesetzte Verweiswörter wie "danach" (7) mit einem einzigem Lemma mit dem Deskriptor *DXZ* erfaßt. Dieses, aus dem Deiktikum "da" und der Präposition "nach" gebildete Wort, das traditionell als Pronominaladverb bezeichnet wird (s. z. B. Helbig & Buscha: 264ff.), wird aufgrund seiner Funktion als zusammengesetztes Verweiswort bestimmt.

Wie die zusammengesetzten Verweiswörter werden auch zusammengesetzte Fragewörter wie "Wie schön, ..." und das türkische Gegenstück "Ne güzel ..." behandelt. Bei "wie" und "ne" handelt es sich isoliert betrachtet um einfache Fragewörter, doch in der Verbindung mit dem folgenden Wort wird die Funktion, den Hörer aufzufordern, in seinem Wissen nach etwas Bekanntem zu suchen, ausgenutzt, um etwas zu suchen, das einer bestimmten Qualität von Schönheit entspricht. Dadurch rückt der ganze Ausdruck in die Nähe der Deixis, er ist paraexpeditiv, insofern er vor allem der Hörerlenkung dient, wobei Mittel des operativen Feldes benutzt werden.

#	Lemma	Wort	W-Art	W-Sem	W-Bld	W-Typ
1	wie	wie	QUE	–	–	QUZ
2	schön	schön	ADJ	–	–	QUZ
3	ne	ne	QUE	–	–	QUZ
4	güzel	güzel	ADJ	–	–	QUZ

Wortübergreifende Ausdrücke: feste Verbindungen mit Verben

Weitere wortübergreifende Ausdrücke sind Infinitivkonstruktionen im Deutschen (s. (1)-(2)), sowie erweiterte Verben, bei denen die Erweiterung mit dem Verb eine Einheit bildet⁹ (s. für das Deutsche (3)-(4), für das Türkische (5)-(6)).

⁹ Die Zusammengehörigkeit zeigt sich u. a. bei der Wortstellung in negierten Äußerungen; s. Aksoy & Grieshaber & Rehbein & Zengin-Kolcu (1992) Lektion 7.

#	Lemma	Wort	W-Art	W-Sem	W-Bld	W-Typ	KST
1	schlafen	schlafen	VER	–	–	–	...
2	gehen	geht	VER	bew	–	–	...
3	Auto	Auto	NOM	–	–	–	...
4	fahren	fahren	VER	bew	–	–	...
5	telefon	telefon	NOM	–	–	–	...
6	etmek	ediyor	VER	–	–	–	...

Die Verbindung wird dadurch erfaßt, daß im Feld KST bei allen beteiligten Einzelwörtern ... eingetragen wird. Aufgrund der sequentiellen Anordnung der Wort-Datensätze können die zusammengehörenden Ausdrücke als Wortfolge bearbeitet werden. In diesen Fällen haben die beteiligten Einzelwörter eine eigene lexikalische Bedeutung, die in der wesentlich syntaktisch begründeten Verbindung erhalten bleibt. Daneben gibt es wortübergreifende Ausdrücke, bei denen die beteiligten Einzelwörter teilweise ihre eigene lexikalische Bedeutung verlieren und in der Bedeutung des gesamten Ausdrucks aufgehen. Dazu zählen Funktionsverbgefüge und idiomatische Ausdrücke. Diese Ausdrücke werden wie erweiterte Verben erfaßt.

§10 Semantische Deskription

Umfang, Struktur und Differenziertheit des Wortschatzes hinsichtlich der Bedeutung stellen ein wesentliches Untersuchungsziel dar. Deshalb werden auch Angaben zur Bedeutung kodiert. Nach dem Grundsatz der wortbezogenen Lemmatisierung und Deskription erfolgt auch die Bedeutungskodierung einzelwortbezogen. Dazu gibt es zu jedem Wort ein entsprechendes Feld. Dieses Verfahren funktioniert problemlos bei Wörtern, die ihrer lexematischen Bedeutung entsprechend verwendet werden. Bei wortübergreifenden Ausdrücken, bei denen die Einzelbedeutung in die Bedeutung des gesamten Ausdrucks eingeht, müssen besondere Verfahren angewendet werden. In diesen Fällen wird bei den Einzelwörtern die lexematische Bedeutung eingetragen, in einem besonderen Feld für wortbezogene Kommentare werden dann Angaben zur Ausdrucksbedeutung erfaßt. Handlungsbezogene Bedeutungsdeskriptoren zur Unterstützung des Situationssamplings können nicht auf der Einzelwortebene ansetzen, sondern auf der Ebene der Segmente. Diese Deskriptoren werden einbeziehen: (a) den Ort einer Handlung oder eines Geschehens, (b) die beteiligten Interaktanten, (c) die durchgeführte Handlung und (d) etwaige benutzte Hilfsmittel zur Durchführung der Handlung.

Die semantische Deskription kann sich, wie oben in §4 gezeigt, nicht auf ein Inventar vorab erstellter Deskriptoren stützen. Deshalb werden zunächst die Verben mit einem groben Raster erfaßt, während die übrigen Symbolfeldausdrücke erst im Verlaufe der Analyse entwickelt werden. Dafür kann dann der listenförmige Zugriff auf ausgewählte Ausdrücke benutzt werden, um die für die Untersuchung geeigneten

Deskriptoren zu gewinnen.

Zu Beginn der Analyse werden zunächst folgende, im Verlauf der Analyse weiter zu entwickelnde, Deskriptoren verwendet:

bew	lexikalische Einheit, die sich auf eine Bewegung bezieht
dic	lexikalische Einheit, die sich auf das Mitteilen usw. bezieht
ext	lexikalische Einheit der Existenz
loc	lexikalische Einheit, die sich auf Ortsbestimmungen bezieht
prs	lexikalische Einheit, die sich auf Personen bezieht
sen	lexikalische Einheit, die sich auf Wahrnehmungen bezieht
tmp	lexikalische Einheit, die sich auf zeitliche Bestimmungen bezieht

Der Schwerpunkt dieser Deskriptoren erfaßt zunächst Verben und - in geringerem Ausmaß - Substantive. Diese beiden Wortarten bilden jedoch bei den vorab durchgeführten Probeläufen mit fünf Kindern den Schwerpunkt des gesamten Lexikons: Verben stellen 31,5%, Substantive 20,4% aller Wörter.

§11 Auswertung

Die Auswertung erfolgt anhand der in dem relationalen Datenbanksystem gespeicherten und beschriebenen Daten. Damit sind die üblichen quantitativen Recherchen zum Type / Token Verhältnis lexikalischer Einheiten realisierbar. Darüber hinaus sind jedoch qualitativ neue Untersuchungsschritte möglich. Es handelt sich um (1) Untersuchungen zur Verwendungsrichtigkeit einzelner Worte, (2) Untersuchungen zur semantischen Struktur des Wortschatzes und (3) Plazierungsverfahren der einzelnen Probanden in Relation zu den in einem Kontext erhaltenen Daten. Im folgenden werden die Untersuchungsschritte kurz beschrieben.

(1) Verwendungsrichtigkeit

In Abhängig von der Erschließbarkeit bietet die KWA eine Analyse der Genauigkeit der Verwendung lexikalischer Einheiten. Bei Bildbeschreibungen kann - u. U. gestützt auf ein Beobachtungsprotokoll - rekonstruiert werden, auf welchen dargestellten Sachverhalt sich eine bestimmte Äußerung bezieht. Damit ist eine Bewertung der Genauigkeit und Richtigkeit der verwendeten sprachlichen Mittel möglich. Dieser Kontextbezug ist nur möglich, wenn die Lemmata jederzeit auf die übergeordneten Segmente bezogen werden können, in denen sie enthalten sind. Bei den üblichen lemmatisierten Wortlisten ist dieser Bezug zur Einzeläußerung nicht mehr gegeben.

In dialogischen Erhebungskontexten kann die Verwendungsrichtigkeit partiell durch Berücksichtigung der interaktionalen Ebene erschlossen werden. Selbst- und Fremverbesserungen sind direkte Ansatzpunkte für die Überprüfung der Genauigkeit der zunächst verwendeten lexikalischen Einheiten. Mitunter kann das Vorliegen eines

Mißverständnisses auch erst nach einer längeren Interaktion manifest werden, wenn einer der Gesprächspartner den aktuellen Gesprächsinhalt nicht mehr in den bis zu diesem Zeitpunkt unterstellten, bzw. verstandenen Verlauf integrieren kann. Die Aufdeckung solcher Mißverständnisse erfordert einen hohen analytischen Aufwand und wird sich deshalb zunächst nur auf interaktional manifeste Probleme beschränken.

In anderen Fällen kann bei entsprechendem Analyseaufwand durchaus auch die Präzision und die formale Richtigkeit der verwendeten lexikalischen Einheiten erschlossen und ausgewertet werden. Bei den schriftlichen Beschreibungen des Tagesablaufs z. B. erlaubt der Vergleich der parallel abgefaßten deutschen und türkischen Versionen die gegenseitige Erhellung der jeweiligen Texte. Dadurch kann im Einzelfall entschieden werden, ob die verwendeten Wörter den gemeinten Sachverhalt treffen, wie folgendes Beispiel zeigt:

(B3) Text-Auszug Mesut Albayrak

deutsch:

- 4) und beim siewesder haben wier Bomben haben wier geblazt

türkisch:

- 3) sonra yılbaşını kutladık.
dann haben wir Neujahr gefeiert.
4) çarşıya Bomba aldık.
auf die Straße haben wir Bomben genommen.

Erst aus dem Vergleich der beiden Versionen ist erschließbar, daß mit "siewesder" keine Schwester, sondern Sylvester gemeint ist und daß 'Bomben blazen' das Zünden von Knallkörpern bedeutet. In diesem Zusammenhang läßt sich entscheiden, daß das für militärische Verwendungen gebräuchliche Wort 'Bombe' im deutschen wie im türkischen Text den gemeinten Sachverhalt nicht treffend wiedergibt, wenn es auch an einer wahrnehmbaren Eigenschaft des Sylvesterknallens ansetzt. Gerade diese Zusammenhänge sollen bei der KWA herausgearbeitet werden.

(2) Wortfelder

Die Untersuchung der semantischen Strukturierung des Wortschatzes umfaßt folgende Einzelaspekte:

- (a) die relative Dichte der Auffächerung eines Wortstammes in verschiedene abgeleitete Wortarten
- (b) die relative Dichte eines Wortfeldes
- (c) die Homogenität der verwendeten lexikalischen Ausdrücke in Bezug auf bestimmte Themen

- (d) die relative Häufigkeit der syntaktischen Einbettung der Wörter in wortübergreifende Ausdrücke.

Ausgangspunkt für die Bestimmung der relativen Auffächerungsdichte sind Wortstämme und die von ihnen abgeleiteten verschiedenen lexikalischen Einheiten. Damit kann ein Maßstab dafür gewonnen werden, über welche Wortbildungstypen ein Sprecher verfügt und wie breit der mit den von ihm benutzten Mitteln Bereich ist.

Die relative Wortfelddichte geht von semantischen Beschreibungen aus und erfaßt die in einem Bereich verwendeten lexikalischen Einheiten. Die Wortfelddichte ergibt ein Maß für die Ausdrucksdifferenzierung.

Die Homogenität der lexikalischen Einheiten innerhalb einzelner Aufgabenstellungen und Themen liefert ebenfalls ein Maß für die Ausdrucksdifferenzierung.

Die Bestimmung der syntaktischen Einbettung der lexikalischen Einheiten ergibt in Verbindung mit der Art des gestellten Aufgabentyps und der davon abhängigen monologischen oder dialogischen Bewältigung ein Maß dafür, ob eher einzelne, unverbunden und isoliert nebeneinander geäußerte Assoziationsnennungen vorliegen oder eine freie Verwendung in syntaktisch vollen Äußerungen erfolgt.

(3) Platzierung relativ zu anderen Schülern

Die Aussagekraft der mit den unter (1) und (2) genannten Analyseverfahren gewonnenen Daten kann durch die Art der Platzierung der Probanden erheblich gesteigert werden. In bisherigen Untersuchungen dominieren lineare Darstellungen, bei denen jeder Proband mittels globaler Prozentangaben für Einzelaufgaben relativ zu den anderen Probanden plazierte wird. Zur Erhöhung der Aussagekraft wird aus den Teilergebnissen ein Globalindex ermittelt. In den jeweiligen Teiluntersuchungen wird die Platzierung nun zusätzlich zur lokalen Platzierung auch vorgenommen in Relation zu den globalen Gesamtergebnissen. Dadurch wird es möglich, Stärken und Schwächen in einzelnen Bereichen der Wortschatzanalysen mit einem globalen Maßstab in Beziehung zu setzen und die Bedeutung des jeweiligen Bereichs für das gesamte Abschneiden zu ermitteln. Nach demselben Platzierungsverfahren werden die Teilergebnisse auch mit den Kontextdaten in Beziehung gesetzt.

Beispiele für Datenrecherchen

Im folgenden werden einige Beispiele für die Auswertung der Daten mittels des relationalen Datenbanksystems Reflex PLUS gegeben, mit dem die KWA zunächst realisiert wird. Durch die Abfragesprache können sämtliche im Dateisystem erfaßten Werte herangezogen werden, auch wenn sie nicht im aktuellen Fenster enthalten sind.

Globale Tokens je Proband

Die generellste Recherche bietet die tabellarische Auflistung der Häufigkeit der einzelnen Wörter. Im folgenden Beispiel sind für die Probanden der Gruppe "OS" je Proband die nach Wortarten aufgeschlüsselten Häufigkeitswerte dargestellt¹⁰:

WortArten global - Gruppe OS - deutsch														
Signle	ADJ	ADV	NOM	VER	VEP	VEA	VZS	DXS	ANA	PRP	WIS	Seg	Wort	
Arz	TR	3	6	46	27	17	15	1	28	1	19	4	52	208
Erd	TR	0	0	10	0	6	7	0	11	0	6	2	24	45
Hsn	TR	3	3	46	5	46	36	2	55	2	14	6	43	252
Met	TR	0	1	5	1	4	2	0	3	0	3	5	5	25
Sun	TR	0	2	21	6	10	10	1	14	2	7	7	13	94
		6	12	128	39	83	70	4	111	5	49	24	137	624

Die Werte in den Zellen der Tabelle können in ein Kalkulationsprogramm exportiert werden und dort als Basis für weitere Berechnungen oder für die Erstellung von Diagrammen dienen.

Differenzierte Type / Token Listen je Proband

Wesentlich differenziertere Einblicke in die Wortschatzstruktur bietet die nach Token strukturierte Auflistung der Lemmata mit den dazugehörigen Wörtern und Segmenten:

Erkan	23	4	163	NOM	Erkana	Erd	dann sind wir nach Erkana gegangen
essen	21	2	151	VEP	gegangen	Erd	um 1 Uhr haben wir Fisch gegessen
fahren	20	1	144	VEP	gefahren	Erd	Ich hab am Samstag um 10 Uhr sind wir nach Markauf gefahren

Der Auszug enthält nur den Kernbereich der Symbolfeldausdrücke (ADJ, ADV, NOM, VER, VEM, VEG, VEP). Die am Ende des Reports aufgelistete type : token Relation von 14 : 16 bestätigt den Eindruck des kurzen Auszugs, daß Erdem bei einem geringen Gesamtumfang nur sehr wenige lexikalische Symbolfeldausdrücke mehrfach verwendet.

¹⁰ Die folgenden Bilder zeigen verkleinerte Ausschnitte aus den Reports des Datenbanksystems.

alisveriş						
124	3	789	NOM	Alışverişe	Erd	Alışverişe gidik
bakmak						
137	16	853	VER	baktık	Erd	saat 20.15 veda tür filmi baktık.
128	7	814	VER	bakma	Erd	Osman 1 Programına bakma istedi
129	8	822	VER	baktık	Erd	Erkanan ikim is yokani kizik 2 Programına baktık
127	6	809	VER	bama	Erd	Emine 2 Programına bama istedi

Im Türkischen ist eine größere lexikalische Variation festzustellen, was durch die type : token Relation von 30 : 51 quantitativ bestätigt wird. Das Verb "bakmak" wird z. B. viermal verwendet und im Report unter dem Lemma "bakmak" in den vier verschiedenen Verwendungsformen aufgelistet. Über die Suchformeln kann der Report umgekehrt auch bei den Wortarten ansetzen und beispielsweise eine Wortart bei allen Probanden auflisten.

Alphabetische Liste, nach Lemmata geordnet:

Die einfachste diagnostische Recherche besteht in der alphabetisch geordneten Liste einzelner Wortarten. Im folgenden Beispiel wurde im File "Worte" nach allen Lemmata gesucht, die im Feld Wortart (wart) den Wert VE enthalten, so daß alle Verben erfaßt werden. Gleichzeitig wird die Suche im Segment-File auf schriftliche deutsche Äußerungen eingeschränkt. Vorab wird die aufsummierte Zahl der mit der Suchformel erfaßten Datensätze angezeigt:

Suchformel: Worte WHERE (INCLUDES(wart,"VE")) AND EXISTS(derSatz WHERE TSpr = "D.s")

Σ gefundene Lemmata: **194**

	wart	wsem	wblid	wkas	wnum	wgen	vtmp	wtyp	KST
604	6:Arz	argern	<input type="checkbox"/>	ergerte	YER				
607	9:Arz	aufhören	<input type="checkbox"/>	aufgehört	YEP				
550	6:Arz	aufstehen	<input type="checkbox"/>	aufgestanden	YEP				

Im folgenden Beispiel wurden die türkischen Verben gesucht und mit derselben Maske wie im obigen Beispiel angezeigt:

Suchformel: Worte WHERE (INCLUDES(wart,"VE")) AND EXISTS(derSatz WHERE TSpr = "TR.s")

Σ gefundene Lemmata: **66**

	wart	wsem	wblid	wkas	wnum	wgen	vtmp	wtyp	KST
692	4:Arz	almak	<input type="checkbox"/>	aldım	YER				
822	7:Erd	bakmak	<input type="checkbox"/>	baktık	YER	sen		-	

Anhand der tabellarischen Auflistung können die Wörter und Lemmata im Überblick und im Kontext der gesamten Äußerung bearbeitet werden, so daß sich diese Darstellung besonders für die Entwicklung und Vereinheitlichung von Deskriptoren eignet.

Das volle Analysepotential des relationalen Datenbankdesigns kann anhand der

kleinen Korpora, die zur Entwicklung des Designs benutzt wurden, noch nicht ausgeschöpft werden. Erst bei erheblich größeren Datenmengen können die wort- oder segmentbezogenen Daten in Relation zu verschiedenen äußeren Kontextbedingungen ausgewählt und differenziert bearbeitet werden.

Schematische Darstellung der Analyseschritte

(1) Datenaufnahme(n):

Mehrere Treffen mit spezifischen Elizitierungsschwerpunkten:

- freie Assoziationen ohne Bezug im aktuellen Handlungskontext;
- Verbalisierung einer Bildvorlage (Einzelbilder, Geschichte);
- Verbalisierung irgendeiner einer bekannten 'Geschichte';
- Verbalisierung irgendeiner erinnerten einzelnen Begebenheit;
- Verbalisierung einer selbsterlebten Handlung (Spielen, Einkaufen).

(2) Verschriftlichung der Daten:

(1) bei mündlichen Äußerungen:

Verlaufsgenaue *Transkription* (HIAT; unter Verwendung von *syncWRITER*)
literarische Transkription; lexikalische Genauigkeit

(2) bei schriftlichen Äußerungen:

genaue *Übertragung* der Texte (evtl. Markierungen für Selbstverbesserungen)

(3) Segmentierung und Export in Datenbank

Segmentierung der Transkripte in Segmentlisten (im Funktionsumfang von *syncWRITER* enthalten)

Segmentierung der Segmentlisten in Wortlisten*

Zuweisung sprachexterner Kontextdaten*

(* unter Verwendung des programmierbaren Editors MEdit und der linguistischen Analyse-Makros LiAM (Grießhaber 1992)

Export der Segment- und Wortlisten in ein relationales Datenbanksystem

(4) Lemmatisierung und Lemmadeskription

Lemmatisierung (unter Beibehaltung der Originalwörter)

Beschreibung der Lemmata mit Deskriptoren

(5) Korpusanalyse

Kontextualisierte Analyse der Daten nach verschiedenen Kriterien

Verwendete Hilfsmittel / Programme:

(1) Reflex PLUS für Macintosh (Borland)

(2) syncWRITER (med-i-bit)

(3) MEdit (Aebi; ShareWare)

Semantische Deskriptoren

Neuner & Schade (1986):

- 1 Personalien
- 2 Wohnen
- 3 Umwelt
- 4 Reisen
- 5 Verpflegung
- 6 Einkaufen / Gebrauchsart.
- 7 öffentliche / priv. Dienstl.
- 8 Gesundheit / Hygiene
- 9 Wahrnehmung / Motorik
- 10 Arbeit / Beruf / Ausbildung
- 11 Freizeit / Unterhaltung
- 12 persönl. Beziehungen
- 13 Aktualität
- 14 Schule
- 15 Metasprache
- 16 Strukturwörter
- 17 neutrale Wörter
- 18 Sonstiges

Textbeispiele deutsch - türkisch**Suzan Aygül**

- 0) Wochenende
 - 1) Samstag morgen waren wir 9.30 in Lübeck in Markauf
 - 2) und hatten etwas gekauft.
 - 3) Dann sind wir 13.30 Uhr Osnabrück zu unsere Bekannten gegangen
 - 4) und haben da Film geguckt.
 - 5) Da nehmen wir noch ein Haus da waren auch unsere Bekannten
 - 6) meine Eltern waren dagewesen.
 - 7) Und wir haben andere Filme geguckt.
 - 8) Und dann kamen meine Vater zurück und sind wieder nach Hause gekommen.
 - 9) Und am Sonntag hatten wir Besuch bekommen und sie hatten zwei Töchter
 - 10) mit ihnen waren wir im Spielplatz und haben da Ball gespielt
 - 11) und sind wieder nach Hause gekommen waren
 - 12) wollten ihre Eltern nach Hause gehen.
 - 13) Und ich habe 18.54 Uhr Fernsehen geguckt dann habe ich geschlafen.
-
- 0) Haftasonu
 - 1) Cumartesi günü sabahlayın 9.30 da Lübeckteki Markafla gıdık
 - 2) ve ordan bıŝeyler aldık ordan sora 13.30 Osnabrücke Misafirliye didik ve orda Film sehtim.
 - 3) Annem gıl onanın evinin üstündekibir ayle daha oturuyordu
 - 4) Annem gıl oraya gidiler.
 - 5) Bıside orda başka Film de sehtik.
 - 6) Sora babamgıl deldiler ve evimize geri döndük.
 - 7) Ve Pazartesi günü bise Misafir geldi ve onların iki tane kısı vardı
 - 8) onlarla parka didik ve Tob oynadık
 - 9) ordan eve gıdiyimıŝ de tam Annesi gıl gıdiyorlardı
 - 10) ve osaman sayt 18.54 geçiyordu onlara yoladık.
 - 11) ben Television baktım ontan sora yatım.

Literaturverzeichnis

- Aksoy, A. & Griebhaber, W. & Rehbein, J. & Zengin-Kolcu, S. (1992) Lehrbuch Deutsch für Türken - Türkler için Almanca ders kitabı. Hamburg: Signum
- Anderson, R. C. & Freebody, P. (1981) Vocabulary knowledge. In: Guthrie, J. T. (ed.) (1981) Comprehension and Teaching: Research Views. Newark: Journal of Reading, 77-117
- Arnaud, P. J. L. (1989) Vocabulary and grammar: A multitrait-multimethod investigation. In: Nation, P. & Carter, R. (eds.) (1989) AILA Review 6/1989, 56-65
- Augst, G. & Bauer, A. & Stein, A. (1977) Grundwortschatz und Ideolekt. Empirische Untersuchungen zur semantischen und lexikalischen Struktur des kindlichen Wortschatzes. Tübingen: Niemeyer
- Augst, G. (1984) Kinderwort. Der aktive Wortschatz kurz vor der Einschulung. Frankfurt/M: Lang
- Augst, G. (Hg.) (1978) Spracherwerb von 6 bis 16. Düsseldorf: Schwann
- Ballmer, T. & Brennenstuhl, W. (1986) Deutsche Verben. Eine sprachanalytische Untersuchung des deutschen Verbwortschatzes.
- Banguoğlu, T. (1986) Türkçenin Grameri. Ankara: Türk Tarih Kurumu Basım Evi
- Benssousan, M. & Laufer, B. (1984) Lexical guessing in context in EFL reading comprehension. In: Journal of Research in Reading 7, 15-32
- Bergenholtz, H. & Schaefer, B. (1977) Die Wortarten des Deutschen. Versuch einer syntaktischen Klassifikation. Stuttgart: Klett
- Broeder, P. & Extra, G. & van Hout, R. & Strömqvist, S. & Voionmaa, K. (eds.) (1988) Processes in the Developing Lexicon. Strasbourg: European Science Foundation
- Broeder, P. & Extra, G. & van Hout, R. (1989) Processes in the developing lexicon of adult immigrant learners. In: Nation, P. & Carter, R. (eds.) (1989) Vocabulary Acquisition. AILA Review 6/89, 86-109
- Brutton, S. R. (1981) An analysis of student and teacher indications of vocabulary difficulty. In: RELC Journal 12, 66-71
- Bühler, K. (1934) Sprachtheorie. Die Darstellungsfunktion der Sprache. Jena: Fischer
- Bühlig, K. (1992) Zur Generalisierung qualitativer Forschungsergebnisse. Fragen der Verknüpfung qualitativer und quantitativer Methoden bei der Untersuchung narrativer Diskursfähigkeiten zweisprachiger Kinder. Hamburg: Germanisches Seminar ENDFAS Arbeitspapier 2 (mimeo)
- Cairns, H. S., Cowart, W. & Jablon, A. D. (1981) Effects of Prior Context upon the Integration of Lexical Information during Sentence Processing. In: Journal of Verbal Learning and Verbal Behaviour 20, 445-453
- Carter, R. & McCarthy, M. (1988) Vocabulary and Language Teaching. London. Longman
- Carter, R. (1987) Vocabulary: Applied Linguistic Perspectives. London: Allen and Unwin
- Clark, E. C. (1973) What's in a word? The child's acquisition of semantics in his first language. In: Moore, T. E. (ed.) (1973) Cognitive Development and the Development of Language. London: Academic Press, 65-110
- Clark, Eve (1974) Some aspects of the conceptual basis for first language acquisition. In: Schiefelbusch, R. L. & Lloyd, L. L. (eds.) (1974) Language Perspectives:

- Acquisition, Retardation and Intervention. New York : Macmillan
- Clark, E. C. (1979) Building a vocabulary: words for objects, actions, and relations. In: Fletcher, P. & Garman, M. (eds.) (1979) Language Acquisition. Studies in First Language Development. Cambridge u.a.: Cambridge University Press, 149-160
- (1981) Lexical innovations: How children learn to create new words. In: Deutsch, W. (ed.) (1981) The Child's Construction of Language. New York: Academic Press, 299-328
- (1982) The young word maker: A case study of innovation in the child's lexicon. In: Wanner, E. & Gleitman, L.R. (eds) (1982) Language Acquisition - The State of the Art. Cambridge : Cambridge University Press, 390-428
- (1983) Convention and contrast in acquiring the lexicon. In: Seiler, Th. & Wannemacher, W. (eds.) (1983) Concept Development and the Development of Word Meaning. Berlin: Springer, 67-89
- (1983) Lexical innovations: How children learn to create new words. In: Deutsch, W. (ed.) (1981) The Child's Construction of Language. London etc.: Academic Press, 299-328
- Clark, E. C. & Clark, H. H. (1977) Psychology and Language. New York: Harcourt Brace Jovanovich
- Clark, E. C. & Frant-Flecht, B. (1982) Learning to coin agent and instrument nouns. In: Cognition 12, 1-24
- Clarke, D. F. & Nation, P. (1980) Guessing the meaning of words from context - strategy and techniques. System 8, 211-220
- Crow, J. (1986) Receptive vocabulary acquisition for reading comprehension. In: Modern Language Journal 70,3, 242-250
- Diersch, H. (1972) Verben der Fortbewegung in der deutschen Sprache der Gegenwart. Berlin: Akademie
- Dollerup, C. & Glahn, E. & Rosenberg Hansen, C. (1989) Vocabularies in the reading process. In: Nation, P. & Carter, R. (eds.) (1989) AILA Review 6/1989, 21-33
- Dornseiff, F. (...) Der deutsche Wortschatz nach Sachgruppen.
- Drewek, R. (1989) LDVLIB - Textanalyse mit System. In: Thaller, M. & Müller, A. (Hg.) (1989) Computer in den Geisteswissenschaften. Frankfurt/M. u. New York: Campus, 9-24
- dtv Wörterbuch der deutschen Sprache. Hrsg. v. G. Wahrig in Zusammenarbeit mit zahlreichen Wissenschaftlern und anderen Fachleuten (1978) München: Deutscher Taschenbuch Verlag
- Duden-Grammatik der deutschen Gegenwartssprache. 4., völlig neu bearb. und erw. Auflage, hrsg. v. G. Drosdowski, Mannheim u.a.: Meyers Lexikonverlag
- Ehlich, K. (1979) Verwendungen der Deixis beim sprachlichen Handeln. Frankfurt/M.: Lang
- (1986) Funktional-pragmatische Kommunikationsanalyse - Ziele und Verfahren. In: Hartung, W.-D. (Hg) (1986) Untersuchungen zur Kommunikation. Berlin: Akademie der Wissenschaften der DDR Zentralinstitut für Sprachwissenschaft. Linguistische Studien Reihe A, Arbeitsberichte 149, S. 15-40-
- (1986) Interjektionen. Tübingen: Niemeyer
- Ehlich, K. & Rehbein, J. (1976) Halbinterpretative Arbeitstranskriptionen (HIAT). In: Linguistische Berichte 45/76, 21-41

- (1977) Wissen, kommunikatives Handeln und die Schule. In: Goeppert, H. (Hg.) (1977) Sprachverhalten im Unterricht. München: Fink, 36-114
- Engel, U. (1988) Deutsche Grammatik. Heidelberg: Groos
- Færch, C. & Kasper, G. (eds.) (1987) Introspection in Second Language Research. Clevedon: Multilingual Matters
- Gerling, M. & Orthen, N. (1979) Deutsche Zustands- und Bewegungsverben: Eine Untersuchung zu ihrer semantischen Struktur und Valenz. Tübingen: Narr
- Ghadessy, M. (1989) The use of vocabulary and collocations in the writing of primary school students in Singapore. In: Nation, P. & Carter, R. (eds.) (1989) AILA Review 6/1989, 110-117
- Goodman, K. & Goodman, Y. & Flores, B. (1984) Reading in the bilingual classroom: literacy biliteracy. Virginia: Inter America Research Associates. National Clearinghouse for Bilingual Education
- Goodman, Kenneth S. & Bridges Bird, Louis (1984) On the wording of texts: A study of intra-text word frequency. In: Research in the Teaching of English 18/1984, 119-145
- Grießhaber, W. (1987) Authentisches und zitierendes Handeln. Band II. Rollenspiele im Sprachunterricht. Tübingen: Narr
- (1988) Geisteswissenschaftliche Arbeitstechniken und Computereinsatz. In: OBST 39/88, 105-128
- (1991) Transkribieren mit *syncWRITER*. Hamburg: Germanisches Seminar (mimeo)
- (1992) LiAM. Linguistische Analysemakros für MEdit 1.3. Hamburg: Germanisches Seminar (mimeo)
- Grundzüge einer deutschen Grammatik. (1981) (Von einem Autorenkollektiv unter der Leitung von Heidolph, K. E. & Flämig, W. & Motsch, W.) Berlin: Akademie
- Haastrup, K. (1989) The learner as word processor. In: Nation, Pa. & Carter, R. (eds.) (1989) AILA Review 6/1989, 34-46
- Hatch, E. (1978) Discourse analysis and second language acquisition. In: Hatch, E. (ed.) (1978) Second Language Acquisition. Rowley, Mass.: Newbury House, 401-435
- Helbig, G. & Buscha, J. (1984) Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht. 8. neubearb. Aufl., Leipzig: Verlag Enzyklopädie
- Hepsöyler, E. & Liebe-Harkort, K. (1988) Wörter und Begriffe. Frankfurt/M.
- Ickler, T. (1987) Paradigmen als Syntagmen. Über das Lernen von Wörtern in Texten. In: FLuL 16/78, 9-24
- Johanson, L. (1990) Studien zur türkeitürkischen Grammatik. In: Hazai, G. (Hg.) (1990) Handbuch der türkischen Sprachwissenschaft. Teil I. Budapest: Akadémiai Kiadó, 146-301
- Johansson, S. (1977) Reading comprehension in the native and the foreign language: On an English-Swedish comprehension index. In: Zettersten, A. (ed.) (1977) Papers on English Language Testing in Scandinavia. Copenhagen: Anglica & Americana 1, 43-58
- Kotsinas, U. (1983) On the acquisition of vocabulary in immigrant Swedish. In: Ringbom, H. (ed.) (1983) Psycholinguistics and Foreign Language Learning. Aebo: Aebo Academi
- Lado, R. (1969²) Moderner Sprachunterricht. München: Hueber

- Laufer, B. (1986) Possible changes in attitude towards vocabulary acquisition research. In: *International Review of Applied linguistics* 24, 69-75
- (1989) A factor of difficulty in vocabulary learning: Deceptive transparency. In: Nation, P. & Carter, R. (eds.) (1989) *AILA Review* 6/1989, 10-20
- Leisi, E. (19612) *Der Wortinhalt. Seine Struktur im Deutschen und Englischen*. Heidelberg: Quelle & Meyer
- Lewis, G. L. (1988) *Turkish Grammar*. Oxford u. New York: Oxford University Press
- Ludwig, J. (1984) Vocabulary acquisition as a function of word characteristics. *The Canadian Modern Language Review* 40, 552-562
- McKeon, M. G. & Curtis, M. E. (eds.) (1987) *The Nature of Vocabulary Acquisition*. Hillsdale, N. J.: Erlbaum
- Meara, P. (1983) *Vocabulary in a Second Language*. Volume 1. (Specialized bibliography 3). London: Centre for Information on Language Teaching and Research (CILT)
- (1987) *Vocabulary in a Second Language*. Volume 2. (Specialized bibliography 4). London: Centre for Information on Language Teaching and Research (CILT)
- (ed.) (1987) Special issue on lexicon. In: *Polyglott* 8
- (1989) Matrix models of vocabulary acquisition. In: Nation, P. & Carter, R. (eds.) (1989) *AILA Review* 6/1989, 66-74
- Meara, P. & Buxton, B. (1987) An alternative to multiple choice vocabulary tests. In: *Language Testing* 4,2, 142-154
- Mrazović, P. & Stölting-Richert, W. (1989) Die Wortschatzbeherrschung zweisprachiger jugoslawischer Schüler in Niedersachsen. In: *Deutsch lernen* 4/89, 28-55
- Nagy, W. E. & Anderson, R. C. (1984) How many words are there in printed school English? In: *Reading Research Quarterly* 20, 233-253
- Nation, P. (1990) *Teaching and Learning Vocabulary*. New York: Newbury House
- Nation, P. & Carter, R. (eds.) (1989) *Vocabulary Acquisition*. In: *AILA Review* 6/1989
- Nation, P. & Coady, J. (1987) *Vocabulary and reading*. In: Carter, R. & McCarthy, M. (eds.) (1987) *Vocabulary and Language Teaching*.
- Nattinger, James R. & DeCarricio, Jeanette S. (1989) Lexical phrases, speech acts and teaching conversation. In: Nation, P. & Carter, R. (eds.) (1989) *AILA Review* 6/1989, 118-139
- Neuner, G. & Schade, E. (1986) *Lernschwierigkeiten ausländischer Schüler mit dem Wortschatz der Grundschule*. Kassel: Kasseler Werkstattberichte zur Didaktik
- Oppertshauer, O. (1974) Absolute oder relative Häufigkeit? Wortstatistik als Hilfsmittel zur Aufstellung eines verbindlichen Mindestwortschatzes für den Englischunterricht im Sekundarbereich I. In: *Praxis des neusprachlichen Unterrichts* 31, 42-52
- Ossner, J. (1989) Wortarten: Form- und Funktionsklassen. Unter besonderer Berücksichtigung ihrer Behandlung in neueren Grammatiken. In: *LiLi* 76/89, 94-117
- Palmberg, R. (1987) Patterns of vocabulary development in foreign-language learners. In: *Studies in Second Language Acquisition* 9,2, 201-219
- (1988) On lexical inferencing and language distance. In: *Journal of Pragmatics* 12,2, 207-214
- (1989) What makes a word English? Swedish speaking learners' feeling of "Englishness". In: Nation, P. & Carter, R. (eds.) (1989) *AILA Review* 6/1989, 47-55
- Politzer, R. (1978) Errors of English speakers of German as perceived and evaluated by German natives. In: *Modern Language Journal* 62/1978, 253-261
- Pregel, D. & Rickheit, G. (1987) *Der Wortschatz im Grundschulalter*. Hildesheim

- Rehbein, J. (1982) Zu begrifflichen Prozeduren in der zweiten Sprache Deutsch. In: Bausch, K.-H. (Hg.) (1982) Mehrsprachigkeit in der Stadtregion. Düsseldorf: Schwann, 225-281
- (1982) Worterklärungen türkischer Kinder. In: OBST 22/82, 122-157
 - (1984) Reparative Handlungsmuster und ihre Verwendung im Fremdsprachenunterricht. Roskilde: Universitetscenter (ROLIG papir 30/84)
 - (1992) Temporalität im Deutschen. Sprachpsychologische und diskursanalytische Betrachtungen zur Tempusverwendung. Hamburg: Germanisches Seminar (mimeo) (Vorläufige Fassung)
 - (1992) Konnektivität komplexer Sätze. Hamburg: Germanisches Seminar. ENDFAS Arbeitspapier 5 (mimeo)
- Richards, J. C. (1970) A psycholinguistic measure of vocabulary selection. In: IRAL 8, 87-102
- Ringbom, H. (1987) The Role of the First Language in Foreign Language Learning. Clevedon: Multilingual Matters
- Schouten-van Parreren, C. (1989) Vocabulary learning through reading. Which conditions should be met when presenting words in texts? In: Nation, P. & Carter, R. (eds.) (1989) AILA Review 6/1989, 75-85
- Special Issue on Vocabulary. (1986) Journal of Reading 7/1986
- Taeschner, T. (1983) The Sun is Feminine. A Study on Language Acquisition in Bilingual Children. Berlin u.a.: Springer
- Taft, M. (1984) Evidence for an abstract lexical representation of word structure. Memory and Cognition 12, 264-269
- Thorndike, E. L. & Lorge, I. (1944) The Teacher's Word Book of 30,000 Words. New York: Columbia University
- Trier, J. (1931) Der deutsche Wortschatz im Sinnbezirk des Verstandes. Heidelberg: Winter
- Van Parreren, C. F. & Schouten-van Parreren, M. Caroline (1981) Contextual guessing, a trainable reader strategy. In: System 9/1981, 235-241
- Vielau, A. (1975) Ausbildung fremdsprachlicher Begriffe. In: Linguistische Berichte 37, 77-91
- (1975) Kognitive Wortschatzarbeit. In: Die Neueren Sprachen 3/1975, 248-264
 - (1977) Der Handlungsaspekt beim Bedeutungslernen. Grundprobleme kognitiver Wortschatzübungen. In: Der fremdsprachliche Unterricht 43/1977, 35-45
- Vihman, M. M. (1981) Phonology and the development of the lexicon: evidence from children's errors. In: Journal of Child Language 8, 239-64
- Vihman, M. M. (1982) Formulas in first and second language acquisition. In: Obler, L. & Menn, L. (ed.) (1982) Exceptional Language and Linguistics. New York : Academic Press, 261-284
- (1982) The acquisition of morphology by a bilingual child: a whole-word approach. In: Applied Psycholinguistics 3, 141-60
- Volterra, V. & Taeschner, T. (1978) The acquisition and development of language by bilingual children. In: Journal of Child Language 5, 311-326
- Zettersten, A. (1979) Experiments in English Vocabulary Testing. Malmö: Liber Hermods
- Zimmermann, H. (1969) Der Aufbau des syntaktischen Wörterbuchs. In: Eggers, H. (Hg.) (1969) Elektronische Syntaxanalyse der deutschen Gegenwartssprache. Tübingen: Niemeyer, 55-61