

# MORPHOLOGIE IN LEXIKALISCHEN DATENBANKEN

## Ein Standardformat zur Repräsentation von morphologischen Eigenschaften lexikalischer Einheiten

Stefan Engelberg

1. Große Lexika mit reichhaltig spezifizierten lexikalischen Einträgen erweisen sich mehr und mehr als Desiderat in der theoretischen wie angewandten Linguistik, insbesondere in fast allen Formen der maschinellen Sprachverarbeitung. Der hohe zeitliche und finanzielle Aufwand der Lexikonerstellung für eine bestimmte Anwendung läßt den Rückgriff auf bereits vorhandene lexikalische Ressourcen wünschenswert erscheinen. Obwohl solche Daten in Form traditioneller und maschinenlesbarer Wörterbücher und Datenbanken vorliegen, ist ihre Auswertung schwierig. Traditionelle Wörterbücher sind aufgrund inhaltlicher Mängel und formaler Inkonsistenzen maschinell nur unter großem Aufwand zu nutzen und in Datenbanken zu konvertieren; lexikalische Datenbanken werden oft auf spezielle Anwendungen zugeschnitten und sind stark vom verwendeten Grammatikformalismus abhängig.<sup>1</sup>

Um diese Schwierigkeiten zu überwinden und die Nutzbarkeit einmal erfaßter lexikalischer Datenbestände zu verbessern, sind in den letzten Jahren verschiedene Projekte ins Leben gerufen worden, die an standardisierten Repräsentationsformaten für lexikalische Daten oder multifunktionalen lexikalischen Datenbanken arbeiten, wie z. B. AQUILEX, GENELEX, LEXIC oder LOLA. Eines dieser Projekte ist MULTILEX, in dessen Rahmen die folgenden Vorschläge am Arbeitsbereich Linguistik in Münster entstanden sind.<sup>2</sup>

2. Der inhaltlichen und formalen Inkompatibilität verschiedener maschinell verfügbarer Lexika soll durch eine Standardisierung der Repräsentation lexikalischer Daten entgegengewirkt werden. Ein solches Standarddatenformat muß verschiedenen Anforderungen genügen:

- (a) Multifunktionalität: Die lexikalischen Informationen sollen für verschiedene Anwendungen im Bereich der maschinellen Sprachverarbeitung und der traditionellen Lexikographie nutzbar sein sowie die Validierung linguistischer Theorien unterstützen.
- (b) Wiederverwendbarkeit: Die Daten sollen in einer Weise formalisiert sein, daß sie in Lexika für verschiedene Anwender konvertierbar sind. (Austauschbarkeit: Die Daten sollen so repräsentiert sein, daß ein Austausch mit vorhandenen lexikalischen Datenbanken möglich ist.)
- (c) Sprachunabhängigkeit: Das Datenformat ist nicht auf eine bestimmte Sprache zugeschnitten; im Rahmen von MULTILEX sollen zunächst zumindest folgende Spra-

chen repräsentierbar sein: Deutsch, Englisch, Französisch, Italienisch, Neugriechisch, Niederländisch, Spanisch.

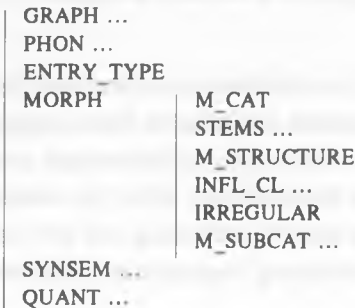
- (d) Empirische Adäquatheit und lexikographische Handhabbarkeit: Das Datenformat soll sowohl die Daten in ihrer ganzen Komplexität empirisch adäquat erfaßbar machen als auch für Lexikographen und Lexikographinnen, die Daten innerhalb dieses Formats erfassen, übersichtlich und handhabbar bleiben.

Als Konsequenz für die inhaltliche Gestaltung des Datenformats ergibt sich eine möglichst große Unabhängigkeit von bestimmten linguistischen Anwendungen und Theorien. Das heißt, daß theoriespezifische Notationen, Begriffe und Generalisierungen vermieden werden sollen. Informationen über sprachliche Phänomene werden unter Verzicht auf verallgemeinernde Klassifizierungen und die Darstellung von Abhängigkeiten zwischen bestimmten Erscheinungen möglichst separat gehalten. Innerhalb eines Lexikonartikels ist daher ein gewisses Maß an Redundanz nicht nur unvermeidbar, sondern auch wünschenswert.

Als Repräsentationsformat dienen Typed Feature Structures, also Argument-Werte-Strukturen, über denen Typen formuliert werden, die in Vererbungshierarchien geordnet sind. In solchen Typen oder Klassen können all die Informationen zusammengefaßt werden, die mehreren lexikalischen Einträgen gemeinsam sind. Durch die Typisierung von Merkmalsstrukturen wird der Anteil vermeidbarer Redundanz gering gehalten und eine hohe Datenintegrität gewährleistet; der Formalismus ist ausreichend mächtig zur Beschreibung verschiedenster Phänomene; zudem sind Typed Feature Structures im Bereich maschineller Sprachverarbeitung und neuer Grammatiktheorien weit verbreitet.<sup>3</sup>

Ein monolingualer Lexikonartikel stellt sich in diesem Format als eine rekursive Attribut-Werte-Matrix dar, wie sie ausschnittsweise in Abb. 1 dargestellt ist.

Abb. 1



Die Struktur ist rekursiv in dem Sinne, daß ein Attribut nicht nur zu einem Wert als terminalem Knoten, sondern auch zu einem weiteren Attribut expandieren kann. Solche Rekursionen erzeugen Pfade, an deren Ende ein atomarer Wert auftritt. So ist in Abb. 2 MORPH | STEMS | BASIC ein Pfad und 'befehl' als Wert zu BASIC atomar.<sup>4</sup>

Die Matrix erlaubt die Spezifizierung folgender Informationen:

- (a) graphemische Information (GRAPH) zu orthographischer Form und Varianz, zur Silbentrennung etc.;
- (b) phonologische Information (PHON) zu Phonemik, Silbenstruktur, Akzent etc.;
- (c) Informationen zum Typ des Eintrags (ENTRY\_TYPE), also ob es sich um ein Lexem, ein phrasales Zeichen, eine Wortform, ein Morphem etc. handelt;
- (d) morphologische Information (MORPH), vgl. Abschnitt 3 und 4;
- (e) syntaktische und semantische Information (SYNSEM) über Wortart und -subklasse, grammatische Kategorien, Valenz, Bedeutung etc.;
- (f) Frequenzinformationen (QUANT) auf der Basis von Korpusauswertungen.

3. Ansätze zur Beschreibung morphologischer Phänomene unterscheiden sich sowohl in der linguistischen Theorie als auch in den Anwendungsbereichen der Lexikographie und der maschinellen Sprachverarbeitung erheblich. In der theoretischen Linguistik konkurrieren etwa wortsyntaktische phrasenstrukturgrammatische Ansätze in der Tradition von Selkirk (1982) und Lieber (1980) mit syntaktisch-transformationellen Ansätzen oder der lexikalischen Phonologie als Interaktion von Phonologie und Morphologie. In der maschinellen Sprachverarbeitung wird lexikalische morphologische Information für verschiedene Analyse- und Syntheseprogramme benötigt.<sup>5</sup> Hier finden sich vor allem zwei unterschiedliche Modelle, die two-level-Morphologie Koskeniemi (1984) und Programme, die auf Unifikationsgrammatiken basieren, sowie Kombinationen daraus (Ritchie et al. 1992, Trost 1990). Zudem besteht auch in der traditionellen Lexikographie wenig Einigkeit über die Beschreibung morphologischer Phänomene, vor allem bei der Behandlung von Wortbildungsregularitäten.

Ein gemeinsamer Nenner für die lexikalische Repräsentation der Morphologie, der allen verschiedenen Theorien und Anwendungen gleichermaßen gerecht wird, ist nur schwer zu finden; und das nicht nur aufgrund theoretischer Divergenzen, sondern auch weil eine Reihe von Phänomenen v. a. im Bereich der Allomorphie bei Wortbildungsprozessen von linguistischer Theorie und Computermorphologie noch nicht erschöpfend behandelt wurden.

Der folgende Vorschlag für die Darstellung morphologischer Eigenschaften im Lexikon kommt den Merkmalskatalogen wortsyntaktischer und unifikationsgrammatischer Ansätze zur Morphologie am nächsten, da diese stark auf deklarativen, merkmalsbasierten Repräsentationen gründen. Im einzelnen liegen folgende Grundsätze der inhaltlichen Gestaltung der Artikelstruktur zugrunde:

- (a) Affixe und Flexive werden in eigenen lexikalischen Einträgen beschrieben;
- (b) lexikographisch kaum erfaßten Phänomenen wird breiter Raum eingeräumt (Stammvarianz in der Derivation, Fugenphänomene, Suffixerweiterungen, Stammerweiterungen);
- (c) konkatenative und nicht-konkatenative Phänomene werden gesondert behandelt; d. h. etwa, daß idiosynkratische Umlaut- und Ablauteigenschaften bei der Flexion getrennt

- von der Flexivsektion (Flexionsklasse) notiert werden, ebenso bei Derivationsprozessen die morphologische Subkategorisierung und die Umlautauslösung;
- (d) Allomorphe bzw. Stammvarianten werden explizit angegeben; auf Generalisierungen durch Regeln wird verzichtet.

4. Das Attribut MORPH in der Matrix (vgl. Abb. 1) erfaßt die morphologischen Eigenschaften von Wortlemmata und Morphemen. Die erforderlichen oder zulässigen untergeordneten Attribute sowie deren Wertebereiche hängen vom Typ des Lemmas, von seiner Wortart und der Einzelsprache ab. Entsprechende Restriktionen über das Kovorkommen eines Attributes mit anderen Attributen oder Werten sind gesondert zu formulieren.

Für das erste Attribut M\_CAT (Morphemkategorie) z. B. gilt als Restriktion, daß es immer und nur dann spezifiziert wird, wenn das Attribut ENTRY\_TYPE den Wert 'morpheme' hat. Weiterhin muß für jedes Attribut dessen Wertebereich beschrieben werden. Der Wertebereich von M\_CAT ist {DERIV\_SUFFIX, INFL\_SUFFIX, PREFIX, FUGE, ENCLITIC, PROCLITIC, SIMPLE\_STEM, VERBAL\_PARTICLE}, ein Morphem ist also entweder ein Derivations- oder Flexionssuffix, ein Präfix, eine Wortbildungsfuge, ein Enklitikon, Proklitikon, Stammorphem oder abtrennbares Verbalpräfix.

Abb. 2

GRAPH	GR_CANON	befehlen		
...				
ENTRY_TYPE	word_lemma			
MORPH	STEMS	BASIC	befehl	
		CHANGED	PRES	befiehl
			PAST	befahl
			PART	befohl
			UMLAUT	beföhl
		JUNCTURES	{befehl}	
	M_STRUCTURE	<befehlen>		
	INFL_CL	strong		
SYNSEM				
...				

Das Attribut STEMS dient zur Notierung von Allomorphie und Stammvarianz, wie sie in Flexion und Wortbildung bei Wörtern und Affixen auftritt. Die atomaren Werte unter STEMS sind Zeichenketten, die den Alloformen des Lemmas entsprechen. BASIC gibt den zugrundeliegenden Stamm des Lemmas an, der in den meisten Fällen mit der kanonischen Zitierform des Lemmas identisch ist. Die Zitierform findet sich in den Beispielen als Wert zu GRAPH | GR\_CANON. Im Deutschen unterscheiden sich Zitierform und zugrundeliegender Stamm bei Verben (z. B. Abb. 2: GR\_CANON | befehlen, BASIC | befehl), nur attributiv verwendeten Adjektiven (z. B. GR\_CANON | obige, BASIC | obig) und Substantiven mit Stammerweiterung (z. B. GR\_CANON | Birne, BASIC | birm).

Abb. 3

GRAPH	GR_CANON	Mann		
...				
ENTRY_TYPE	word_lemma			
MORPH	STEMS	BASIC	mann	
		CHANGED	UMLAUT	männ
		JUNCTURES	{mann, manns, männer, mannes}	
	M_STRUCTURE	< mann >		
	INFL_CL	SG	1	
		PL	4	
		PL_UMLAUT	+	
SYNSEM				
...				

Mit dem Attribut CHANGED werden Stammalternanzen in Flexions- und Derivationsprozessen erfaßt; seine untergeordneten Attribute hängen ab von der Einzelsprache, der Wortart und der Flexionsklasse des Lemmas. Bei deutschen Substantiven können hier die Attribute UMLAUT und EXTENDED auftreten. Wie alle nicht-konkatenativ erzeugten Formen wird auch die Umlautform explizit angegeben (z. B. Abb. 3).

Mit dem Attribut EXTENDED wird das Phänomen der Stammerweiterung beschrieben (vgl. Abb. 4). Bei Substantiven wie *Birne*, *Riemen*, *Epos*, *Liga* treten einige Derivations- und Flexionssuffixe an einen um eine Stammerweiterung verkürzten Stamm: *Birn-chen*, *Riem-chen*, *ep-isch*, *Ep-en*, *Lig-en*. Das Attribut EXTENDED notiert das Wort mit Stammerweiterung, unter BASIC findet sich der Stamm ohne Stammerweiterung.<sup>6</sup>

Abb. 4

GRAPH	GR_CANON	Epos		
...				
ENTRY_TYPE	word_lemma			
MORPH	DOC_FFORMS	...		
	STEMS	BASIC	ep	
		CHANGED	EXTENDED	epos
		JUNCTURES	{epos, epen}	
	M_STRUCTURE	< epos >		
	INFL_CL	SG	3	
		PL	2	
		PL_UMLAUT	-	
SYNSEM				
...				

Die anderen Attribute unter CHANGED spezifizieren v. a. die verschiedenen für die Konjugation erforderlichen Verbalstämme, in Abb. 2 etwa den Stamm für die zweite und dritte Person Präsens Singular (PRES), für das Präteritum (PAST), das Partizip II (PART) und den Konjunktiv Präteritum (KONJ). Bei spanischen Verben (vgl. Abb. 5) können zwei Alternanzstämme auftreten: PRES z. B. für den Singular und die dritte Person des Präsens Indikativ, PAST u. a. für die dritte Person Singular und Plural des unbestimmten Präteritums.

Abb. 5

GRAPH	GR_CANON	consentir			(= zustimmen)
...					
ENTRY_TYPE	word_lemma				
MORPH	STEMS	BASIC	consent		
		CHANGED	PRES	consient	
			PAST	consint	
	M_STRUCTURE	< con, sentir >			
	INFL_CL	regular			
SYNSEM					
...					

Unter JUNCTURES wird die Form eines Lemmas innerhalb eines Kompositums notiert. Die Form, die ein Wort als erstes Glied in einem binären Kompositum annimmt, ist im Deutschen eine weitgehend idiosynkratische Eigenschaft dieses Wortes. Das Substantiv *Mann* (vgl. Abb. 3) kann in Komposita in den Formen *mann*, *manns*, *männer*, *mannes* auftreten (*Mannjahr*, *Mannsbild*, *Männerklo*, *Manneskraft*). Das Verb *sehen* findet sich in Komposita als *seh* (*Sehkraft*, *Sehhilfe*), *lesen* als *lese* (*Lesebuch*, *Lesevergnügen*).<sup>7</sup>

Das nächste Attribut M\_STRUCTURE gibt die morphologische Struktur des Lemmas an als Liste der Einheiten, aus denen es besteht. Die Einheiten werden nicht als Allomorphe, sondern immer als Zitierformen notiert. Der Wert für *Arbeitsamt* wäre also nicht <arbeits, amt> oder <arbeit,s,amt>, sondern <Arbeit,Amt>, der Wert für *Lädchen* <Laden, chen> und nicht <läd,chen>. Der Zugriff auf die Artikel zu den Konstituenten komplexer Wörter wird so erleichtert; sowohl die Allomorphie der Konstituenten als auch die grammatischen Eigenschaften des komplexen Wortes lassen sich über einschlägige Head- und Perkolationsmechanismen aus den Informationen zu den Konstituenten vorhersagen. Verzichtet wird auf die Angabe der Konstituentenstruktur, die stark von bestimmten theoretischen Entscheidungen abhängig ist und besser über Regeln erfaßt wird.<sup>8</sup>

INFL\_CL enthält die Flexionsklasse des Lemmas je nach Sprache und Wortart mit entsprechenden weiteren Attributen, für deutsche Substantive etwa SG für die Flexionsklasse im Singular und PL für die Pluralflexionsklasse, sowie PL\_UMLAUT als binäres Merkmal zur Kennzeichnung, ob das Substantiv im Plural in der umgelauteten Variante auftritt (vgl. Abb. 3, 4). Flexionsklasse wird hier verstanden als die Eigenschaft, bestimmte Flexive zu selektieren. Nicht-konkatenativ entstandene Flexionsformen werden in MORPH | STEMS | CHANGED bzw. IRREGULAR erfaßt.<sup>9</sup>

Das Merkmal IRREGULAR enthält als Wert die Menge der flexivischen Suppletivformen zu einem Lemma, also alle die Formen, die nicht über die regulären Flexionsmerkmale erzeugt werden können. Die Suppletivformen sind selbst lexikalische Einträge (ENTRY\_TYPE | inflected\_form) und werden dort bezüglich ihrer grammatischen Kategorien spezifiziert.

Abb. 6

GRAPH	GR_CANON	-er	(z. B. Frager, Jäger)
...			
ENTRY_TYPE	morpheme		
MORPH	M_CAT	deriv_suffix	
	STEMS	BASIC	-er
		JUNCTURES	{-er,-ers}
	M_STRUCTURE	<er>	
	INFL_CL	SG	1
		PL	3
		PL_UMLAUT	-
	M_SUBCAT	BASE_CAT	{v}
		BASE_STEM	{basic,umlaut}
		ACCENT_TYPE	stem_accent
SYNSEM	HEAD	PART_OF SPEECH	n
		SUBCLASS	count
		GENDER	masc
...			

Das Merkmal M\_SUBCAT (morphologische Subkategorisierung) kommt nur Affixen und Verbalpartikeln zu. Es beschreibt ihre Selektionseigenschaften als Klasse von Wörtern, an die sie affigieren können (vgl. Abb. 6). BASE\_CAT gibt dabei die Wortart der Derivationsbasis an; das Suffix *-er* etwa affigiert nur an Verben. BASE\_STEM spezifiziert die Stammvariante, an die das Affix tritt; Neubildungen mit *-er* treten entweder an den zugrundeliegenden oder den umgelauteten Stamm.

Abhängig von dem Affix bleibt der Akzent einer Ableitung entweder auf dem Stamm, an den das Affix tritt, z. B. im Englischen *cóncscious* --> *cóncsciousness* (ACCENT\_TYPE | stem\_accent), oder auf dem Affix selbst, z. B. *Búchstabe* --> *buchstabieren* (AFFIX\_TYPE | affix\_accent). Andere Affixe wiederum bewirken, daß der Akzent durch alle Flexionsformen hindurch auf der gleichen Silbenposition bleibt, bei *-or* etwa immer auf der vorletzten Silbe: *Reflektor*, *Reflektóren*.<sup>10</sup>

Andere Merkmale komplexer Wörter, die durch das Affix bestimmt werden, wie etwa Wortart oder Genus finden sich unter dem Attribut SYNSEM. Das Suffix *-er* z. B. bildet zählbare, maskuline Substantive.<sup>11</sup>

## Anmerkungen

- 1 Vgl. dazu etwa Calzolari (1989), Van der Eijk et al. (1992).
- 2 MULTILEX ist ein Gemeinschaftsunternehmen verschiedener europäischer Universitäten, Forschungsinstitute und mit Softwareentwicklung befaßter Unternehmen. Im Rahmen von MULTILEX wird ein Standard für ein multifunktionales mehrsprachiges Lexikon entwickelt, zusammen mit Werkzeugen zur Verwaltung und Nutzung dieser Daten. Eine lexikale Datenbank für mehrere Sprachen entsteht in diesem Format und soll in verschiedenen Anwendungsbereichen getestet werden.
- 3 Zu Typed Feature Structures und einigen Problemen mit multipler Vererbung und mit der Repräsentation

- bestimmter Informationstypen vgl. Fedder (1992).
- 4 Die Pünktchen in der Matrix markieren ausgelassene Attributpfade.
  - 5 Einen Überblick über den Einsatz morphologischer Programme und Programmkomponenten geben Schaefer/Willée (1989).
  - 6 Ein ähnliches Phänomen findet sich im Neugriechischen, wo einige Substantive im Plural, z. T. auch im Genitiv Singular eine Stammerweiterung *-d-* oder *-t-* vor dem Flexiv zeigen, z. B. Nominativ Singular *psomá-s* (Bäcker), Nominativ Plural *psomá-d-es*.
  - 7 Über die Schwierigkeit der Analyse deutscher Komposita und den Mangel an Informationen in maschinenlesbaren Wörterbüchern zur Allomorphie von Substantiven in Komposita vgl. Dagan et al. (1992).
  - 8 Probleme ergeben sich hier bezüglich der Behandlung von Klammerparadoxien, der Annahme einer Binärstruktur komplexer Wörter oder der Auflösung von Zirkumfigierungen.
  - 9 Die Trennung konkatenativer von nicht-konkatenativen Flexionsphänomenen sowie Singular- von Pluralflexionsklassen entspricht unserem Prinzip, Informationen möglichst separat zu halten. Im Spanischen etwa ist für Verben zwar die Angabe ablautähnlicher Stammalternanzen erforderlich, die Flexivselektion ist aber völlig regulär. Ebenso sollten Singular- und Pluralflexion bei deutschen Substantiven als unabhängige Phänomene notiert werden. Weder ist die Pluralflexionsklasse immer eindeutig aus der Singularflexionsklasse ableitbar noch umgekehrt.
  - 10 Zwei weitere Merkmale, wie sie etwa von Ritchie et al. (1992: 125, 215) für ihren morphologischen Parser verwendet werden, stehen zur Diskussion. Einige Suffixe im Englischen (*-ion*, *-ive*, *-ory*) haben eine Variante mit der Suffixerweiterung *-at-* (*-ation*, *-ative*, *-atory*). Stämme, die sich mit diesen Suffixen verbinden, wählen entweder immer die erste Variante oder die zweite und werden entsprechend als AT- oder AT+ spezifiziert. Die Suffixe wiederum sind entsprechend dafür markiert, daß sie für Stämme des Typs AT- oder AT+ subkategorisieren. Des weiteren sollen Stämme gekennzeichnet werden, ob sie lateinischen Ursprungs sind oder nicht (LAT-, LAT+). Dem liegt die Beobachtung zugrunde, daß bestimmte Affixe für Stämme lateinischen Ursprungs subkategorisieren, andere für Stämme nichtlateinischen Ursprungs. Die Eigenschaften bestimmter Affixe, an Stämme des Typs LAT: + zu affigieren, den Akzent seiner Basis zu verändern, über phonetisch bedingte Allomorphie zu verfügen und näher am Wortstamm zu stehen als Affixe des gegenteiligen Typs, korrelieren in hohem Maße. Das führte zur Annahme bestimmter Affixklassen (vgl. fürs Englische Siegel (1974), fürs Deutsche Giegerich (1985)). Gemäß dem Grundsatz, auf theoriebedingte Klassifizierungen zu verzichten, werden diese Informationen separat in verschiedenen Merkmalen erfaßt: `PREFERRED_STEM`, `ACCENT_TYPE`, evtl. `LAT`.
  - 11 Hier sollen in Abstimmung mit der Bedeutungsrepräsentation von lexikalischen Einheiten auch die valenzverändernden Eigenschaften von Affixen beschrieben werden.

## Literatur

- Bátori, I. / Lenders, W. / Putschke, W. (eds.) (1989): Computational linguistics. Computerlinguistik. Ein internationales Handbuch zur computergestützten Sprachforschung und ihrer Anwendungen. - Berlin etc.: de Gruyter.
- Calzolari, N. (1989): "Computer-aided lexicography: Dictionaries and word data bases". - In: Bátori / Lenders / Putschke (eds.) 510-519.
- COLING-92 (1992). Proceedings of the fifteenth International Conference on Computational Linguistics. I-IV. - Nantes.
- Dagan, I. / Rackow, U. / Schwall, U. (1992): "Automatic translation of noun compounds". - In: COLING-92 (1992) 1249-1253.
- Eijk, P. van der / Bloksma, L. / Van der Kraan, M. (1992): "Towards developing reusable NLP dictionaries". - In: COLING-92 (1992) 53-59.
- Fedder, L. (1992): Formalism. MULTILEX internal format. - MULTILEX internal report (unveröffentl.).
- Giegerich, H. J. (1985): Metrical phonology and phonological structure. - Cambridge etc.: CUP.



- Koskenniemi, K. (1984): "A general computational model for word-form recognition and production". - In: COLING-84 (1984). Proceedings of the tenth International Conference on Computational Linguistics.
- Lieber, R. (1980): The organization of the lexicon. - Cambridge, Mass.: MIT (vervielf.).
- Ritchie, G. D. / Russell, G. J. / Black, A. W. / Pulman, S. G. (1992): Computational morphology. Practical mechanisms for the English lexicon. - Cambridge, Mass. etc.: MIT.
- Schaeder, B./ Willée, G. (1989): "Computergestützte Verfahren morphologischer Beschreibung". - In: Bátori / Lenders / Putschke (eds.) 188-203.
- Selkirk, E. (1982): The Syntax of words. - Cambridge, Mass.: MIT.
- Siegel, D. (1974): Topics in English morphology. - New York etc.: MIT.
- Trost, H. (1990): Recognition and generation of word forms for natural language understanding systems: Integrating two-level-morphology and feature unification. - Saarbrücken: Deutsches Forschungszentrum für Künstliche Intelligenz.