

# Contents/Inhalt

Preface.....	XIII
Vorwort .....	XXI
List of Maps/Verzeichnis der Karten .....	XXIX
List of Figures/Verzeichnis der Abbildungen .....	XXX
<b>I. Computational Linguistics I : Development</b> <b>Computerlinguistik I : Entwicklungslinien</b>	
1. István S. Bátori, Die allgemeine Entwicklung der Computerlinguistik ( <i>The General Development of Computational Linguistics</i> ).....	1
2. Brigitte Endres-Niggemeyer, Überblick über die Wissenschaftsorganisa- tion und Forschungseinrichtungen der Computerlinguistik in den westli- chen Ländern ( <i>A Survey of the Scientific Organization and Research Fa- cilities of Computational Linguistics in the Western Countries</i> ) .....	14
3. Eva Hajičová, The Research in Computational Linguistics in the Coun- tries of Eastern (and Central) Europe ( <i>Forschungen zur Computerlingui- stik in Ost-[und Mittel-]Europa</i> ) .....	59
<b>II. Computational Linguistics II: Theoretical and Methodical</b> <b>Foundations</b> <b>Computerlinguistik II: Theoretisch-methodische Grundlagen</b>	
4. Willy Martin, Theoretical Framework, Aims and Concepts of Computa- tional Linguistics ( <i>Wissenschaftstheoretische Grundlagen, Ziele und Kon- zeptionen der Computerlinguistik</i> ) .....	73
5. Dieter Krallmann, Problembereiche der Computerlinguistik: Positionie- rungs- und Abgrenzungsaspekte ( <i>Problems of Computational Linguistics: Its Relation to Other Fields of Research</i> ) .....	79
6. Ursula Klenk, Computerlinguistik und die Theorie der formalen Spra- chen ( <i>Computational Linguistics and the Theory of Formal Languages</i> ).....	87
7. Eberhard Bertsch, Abbildungsproblematik in der Computerlinguistik: Linguistische Elemente und Operationen vs. Daten- und Programmstruk- turen ( <i>The Problem of Mapping in Computational Linguistics: Linguistic Elements and Operations vs. Data- and Program Structures</i> ) .....	94
8. Hans Karlgren, Application and Research — Mutual Side Effects ( <i>An- wendung und Forschung — Wechselseitige Nebeneffekte</i> ) .....	101
<b>III. Computational Linguistics III: Special Methodical Problems</b> <b>Computerlinguistik III: Besondere methodische Probleme</b>	
9. Reinhard Köhler/Gabriel Altmann, Status und Funktion quantitativer Verfahren in der Computerlinguistik ( <i>Status and Function of Quantitative Methods in Computational Linguistics</i> ) .....	113

10.	Jürgen Krause, Evaluierung komplexer Systeme in der Computerlinguistik: Verfahren und Anwendung ( <i>The Evaluation of Complex Systems in Computational Linguistics: Methods and Applications</i> ) .....	120
IV. Computer-Aided Description of Language Systems I: Basic Problems Computergestützte Beschreibung von Sprache I: Grundprobleme		
11.	Petr Sgall, Theoretical Aspects of Computer-Aided Description of Language Systems ( <i>Theoretisch-methodische Aspekte der computergestützten Beschreibung von Sprache</i> ) .....	127
12.	Henning Bergenholtz/Joachim Mugdan, Korpusproblematik in der Computerlinguistik: Konstruktionsprinzipien und Repräsentativität ( <i>Problems of Natural Language Corpora in Computational Linguistics: Principles of Construction and Representativity</i> ) .....	141
13.	Robert F. Allen, Survey of Natural Language Corpora in Computational Linguistics ( <i>Übersicht über Korpora in..... der Computerlinguistik</i> ).....	150
14.	Winfried Lenders, Segmentierung in der Computerlinguistik ( <i>Segmentation in Computational Linguistics</i> ) .....	159
V. Computer-Aided Description of Language Systems II: Discovery of Rule-Systems Computergestützte Beschreibung von Sprache II: Ermittlung von Regelsystemen		
15.	Ilpo Tapani Piirainen, Computergestützte Verfahren zur graphematischen Beschreibung von Sprache ( <i>Computer-Aided Graphematic Description of Language Systems</i> ) .....	167
16.	Klaus Wothke, Computergestützte Verfahren zur phonologischen Beschreibung ( <i>Computer-Aided Phonological Description</i> ).....	175
17.	Burkhard Schaefer/Gerd Willée, Computergestützte Verfahren morphologischer Beschreibung ( <i>Computer-Aided Morphological Description</i> ) .....	188
18.	Allan Ramsay, Computer-Aided Syntactic Description of Language Systems ( <i>Computergestützte Verfahren zur syntaktischen Beschreibung von Sprache</i> ) .....	204
19.	Hans-Jürgen Eikmeyer/Jörg Kindermann/János Sándor Petőfi, Computer-Aided Textual Description of Language Systems ( <i>Computergestützte Verfahren zur textualen Beschreibung von Sprache</i> ) .....	219
20.	Winfried Lenders, Computergestützte Verfahren zur semantischen Beschreibung von Sprache ( <i>Computer-Aided Semantic Description of Language Systems</i> ) .....	231

<b>VI.</b>	<b>Computer-Aided Description of Language Systems III: Testing of Rule Systems Computergestützte Beschreibung von Sprache III: Testen von Regelsystemen</b>	
21.	Fred Karlsson, Computational Testing of Linguistic Models in Morphology ( <i>Computergestützte Testverfahren für Sprachbeschreibungsmodelle im Bereich der Morphologie</i> ) .....	245
22.	Joyce B. Friedman, Computational Testing of Linguistic Models in Syntax and Semantics ( <i>Computergestützte Testverfahren für Sprachbeschreibungsmodelle im Bereich der Syntax und Semantik</i> ) .....	252
<b>VII.</b>	<b>Computer Simulation of Language I: Basic Problems Computersimulation von Sprache I: Grundprobleme</b>	
23.	Winfried Lenders, Übersicht über die Verstehensproblematik hinsichtlich der Computersimulation von Sprache ( <i>Survey of the Problem of Understanding from the Point of View of Computer Simulation of Language</i> ) .....	260
24.	Gerry T. M. Altmann, Psycholinguistic Implications in Computer Simulation of Natural Language Processing ( <i>Psycholinguistische Implikationen der Computersimulation von Sprachprozessen</i> ) .....	272
25.	Haldur Öim, Language Understanding and Problem Solving: On the Relation Between Computational Linguistics and Artificial Intelligence ( <i>Sprachverstehen und Problemlösen: Zum Verhältnis von Computerlinguistik und Künstlicher Intelligenz</i> ) .....	277
26.	Geneviève Berry-Rogghe, Representational Language in Language Understanding Systems ( <i>Repräsentationssprache in Sprachverstehenssystemen</i> ) .....	283
27.	Dieter Metzging/Günther Görz, Rolle und Funktion von Netzwerken in der Computersimulation von Sprache ( <i>The Use of Networks in Computer Simulation of Language</i> ) .....	290
<b>VIII.</b>	<b>Computer Simulation of Language II: Systems Components Computersimulation von Sprache II: Systemkomponenten</b>	
28.	István S. Bátori, Systemarchitektur der Sprachverstehenssysteme ( <i>The Architecture of Language Understanding Systems</i> ) .....	311
29.	Manfred Wettler, Wissenrepräsentation: Typen und Modelle ( <i>Knowledge Representation: Types and Models</i> ) .....	317
30.	Robert J. P. Ingria, Simulation of Language Understanding: Lexical Recognition ( <i>Simulation sprachlichen Verstehens: Lexikalisches Erkennen</i> ) ...	336
31.	Peter Hellwig, Parsing natürlicher Sprachen: Grundlagen ( <i>Parsing of Natural Language Systems: Foundations</i> ) .....	348
32.	Peter Hellwig, Parsing natürlicher Sprachen: Realisierungen ( <i>Parsing of Natural Language Systems: Realisations</i> ) .....	378

33.	Claus-Rainer Rollinger, Simulation sprachlichen Verstehens: Generelle Probleme bei der semantischen Interpretation der natürlichen Sprache ( <i>Simulation of Language Understanding: General Problems in Semantic Interpretation of Natural Language</i> ) .....	432
34.	Margaret King, Simulation of Language Understanding: Semantic Models and Semantic Interpretation ( <i>Simulation sprachlichen Verstehens: Semantische Modelle und semantische Interpretation</i> ) .....	440
35.	Hans Dieter Lutz, Inferenzen in sprachverstehenden Systemen ( <i>Inferences in Language Understanding Systems</i> ) .....	460
36.	Gerard Kempen, Language Generation Systems ( <i>Spracherzeugungssysteme</i> ) .....	471
 IX. Application Fields of Computational Linguistics I: Basic Problems Anwendungsgebiete der Computerlinguistik I: Grundprobleme		
37.	István S. Bátori, Grundprobleme der Anwendungen in der Computerlinguistik ( <i>Basic Problems of Applications in Computational Linguistics</i> ) .....	481
 X. Application Fields of Computational Linguistics II: State of the Art and Perspectives of Computer-Aided Description of Natural Language Anwendungsgebiete der Computerlinguistik II: Stand und Perspektiven der computergestützten Beschreibung natürlicher Sprache		
38.	Randall L. Jones/Steven P. Sondrup, Computer-Aided Lexicography: Indexes and Concordances ( <i>Computergestützte Lexikographie: Indices und Konkordanzen</i> ) .....	490
39.	Nicoletta Calzolari, Computer-Aided Lexicography: Dictionaries and Word Data Bases ( <i>Computergestützte Lexikographie: Wörterbücher und Wortdatenbanken</i> ) .....	510
40.	Klaus-Peter Wegera/Eckhard Berg, Computergestützte Grammatikographie. Eine Fallstudie ( <i>Computer-Aided Grammaticography: A Case Study</i> ) ....	519
41.	Dieter Wickmann, Computergestützte Philologie: Bestimmung der Echtheit und Datierung von Texten ( <i>Computer-Aided Philology: Authorship and Chronological Determination</i> ) .....	528
42.	Dietmar Najock, Computer-Assisted Reconstruction of Texts ( <i>Computergestützte Rekonstruktion von Texten</i> ) .....	534
43.	Robert F. Allen, Computer-Aided Stylistic Analysis. A Case Study of French Texts ( <i>Computergestützte Stilanalyse. Eine Fallstudie zu französischen Texten</i> ) .....	544
44.	Harald Händler/Lutz Hummel/Wolfgang Putschke, Computergestützte Dialektologie ( <i>Computer-Aided Dialectology</i> ).....	553

45.	John Hewson, Computer-Aided Research in Comparative and Historical Linguistics ( <i>Computergestützte Untersuchungen in der vergleichenden und historischen Linguistik</i> ) .....	576
46.	Peter Ph. Mohler, Computergestützte Inhaltsanalyse: Überblick über die linguistischen Leistungen ( <i>Computer-Aided Content-Analysis: A Survey of Linguistic Contributions</i> ) .....	580
XI.	Application Fields of Computational Linguistics III: State of the Art and Perspectives in Computer-Aided Simulation of Natural Language in Man-Machine-Communication Anwendungsgebiete der Computerlinguistik III: Stand und Perspektiven der Computersimulation natürlicher Sprache in der Mensch-Maschine-Kommunikation	
47.	Hermann Ney, Automatic Speech Recognition ( <i>Automatische Sprach-erkennung</i> ) .....	586
48.	Bernd S. Müller, Automatische Sprachsynthese ( <i>Automatic Speech Synthesis</i> ) .....	599
49.	Alan Melby, Machine Translation: General Development ( <i>Maschinelle Sprachübersetzung: Allgemeine Entwicklung</i> ) .....	622
50.	Jonathan Slocum, Machine Translation: A Survey of Active Systems ( <i>Maschinelle Sprachübersetzung: Überblick über aktive Systeme</i> ) .....	629
51.	Károly Fábircz, Machine Translation: Active Systems. East-European Countries ( <i>Maschinelle Sprachübersetzung: Aktive Systeme. Osteuropa</i> ) ..	645
52.	Jun-ichi Tsujii, Machine Translation: Research and Trends ( <i>Maschinelle Sprachübersetzung: Forschungstrends</i> ) .....	652
53.	Christian Boitet, Software and Lingware Engineering in Modern M(A)T Systems ( <i>Software und Lingware-Technik in modernen M(A)T-Systemen</i> ) .....	670
54.	Yu. M. Marchuk, Machine-Aided Translation: A Survey of Current Systems ( <i>Maschinengestützte Sprachübersetzung: Überblick über Systeme</i> ) ..	682
55.	Rainer Kuhlen, Information Retrieval: Verfahren des Abstracting ( <i>Information Retrieval: Methods in Abstracting</i> ) .....	688
56.	Jiri Panyr/Harald H. Zimmermann, Information Retrieval: Überblick über aktive Systeme und Entwicklungstendenzen ( <i>Information Retrieval: A Survey of Active Systems and Trends</i> ) .....	696
57.	Christopher Habel/Simone Pribbenow, Frage-Antwort-Systeme: Übersicht über Konzeptionen, Leistungsumfang und Entwicklungsperspektiven ( <i>Question-Answering-Systems: A Survey of Models, Performance and Trends</i> ) .....	708
58.	Arkadiusz Lesch, Question-Answering-Systems: A Survey of Major Accomplishments ( <i>Frage-Antwort-Systeme: Überblick über die Leistungen</i> ) .....	722
59.	Bruce W. Ballard, Computer-Aided Office Automation: A Survey of Linguistic Contributions ( <i>Computergestützte Bürokommunikation: Überblick über die linguistischen Leistungen</i> ) .....	739
60.	Arne Zettersten, Computer-Assisted Instruction: Problems and Perspectives from the Point of View of Computational Linguistic ( <i>Computergestützter Unterricht: Probleme und Möglichkeiten aus der Sicht der Computerlinguistik</i> ) .....	745

## XII. Operating Environment of Computational Linguistics Das arbeitspraktische Umfeld der Computerlinguistik

- |     |   |     |
|-----|---|-----|
| 61. | Susan Hockey, A Survey of Practical Aspects of Computer-Aided Maintenance and Processing of Natural Language Data ( <i>Überblick über arbeitspraktische Aspekte der computergestützten Haltung und Bearbeitung natürlich-sprachlicher Daten</i> ) .....     | 752 |
| 62. | Raimund Drewek, Arbeitspraktische Probleme des Datenaustauschs: Normungstendenzen und Kompatibilität im Bereich linguistischer Daten ( <i>Practical Problems of Data Exchange: Standards and Compatibility of Data in Computational Linguistics</i> ) ..... | 760 |
| 63. | Jan Messerschmidt/Günter Hotz, Eignung von Programmiersprachen zur Lösung linguistischer Problemstellungen: Entwicklung und Perspektiven ( <i>Capacity of Programming Languages for Linguistic Problem Solving: Development and Trends</i> ) .....          | 766 |
| 64. | Jonko Seppänen, Special Hardware and Future Development in Computer Architecture in View of Computational Linguistics ( <i>Spezielle Hardware und Entwicklungen auf dem Gebiet der Computerarchitektur aus der Sicht der Computerlinguistik</i> ) .....     | 772 |
| 65. | Harald Händler, Graphikumgebung als arbeitspraktisches Ausstattungsbeispiel ( <i>Operating Graphical Facilities in Computational Linguistic: From a User's Point of View</i> ) .....  | 790 |

## XIII. Bibliography Bibliographie 799

## XIV. Indexes Register

- |     |  |     |
|-----|--|-----|
| 66. | Index of Names ( <i>Namenregister</i> )..... | 901 |
| 67. | Subject Index ( <i>Sachregister</i> ) .....  | 917 |

## Preface

### 1. On the Term “Computational Linguistics”

During the last four decades, the subject called *Computational Linguistics* in this handbook has evolved from humblest beginnings. In the practical context of research, tuition, and application, the terms *Linguistische Datenverarbeitung*, *Computative Linguistik*, *Sprachdatenverarbeitung*, *Linguistische Informatik*, *Informationslinguistik*, *literary and linguistic computing*, *language data processing*, etc., have also been used, referring to a more extensive field that covers those sections of artificial intelligence research that are related to language as well. The various names for the subject reflect the interests of disciplines such as linguistics, philologies, information science, psychology, etc., as well as some areas of application, like language translating, information systems, text processing, etc. In spite of all its different names, this subject deals with a clearly defined topic: language and language phenomena, the research of which is carried out with the special aid of electronic data processing. The title of the handbook has been chosen thus because, in the editors' opinion, the term *Computational Linguistics* seems to be most suitable to express the intention of bringing together differently accentuated fields of work. Moreover, *Computational Linguistics* has already been adopted as the name of the largest international association of specialists in this field of work.

### 2. On the Origin of this Handbook

It was the late co-editor of the series “Handbücher zur Sprach- und Kommunikationswissenschaft”, Gerold Ungeheuer, who was first to recognize the need for this handbook, and who committed himself to the all important inclusion in the series. The editors dedicate this volume to his memory.

The preliminary talks took place in 1981. Introductions for the handbook and an outline of its structure were elaborated, distributed and discussed with colleagues in the Computational Linguistics Conferences in Prague (1982) and Palo Alto/Stanford (1984). The editors did not only note a mood of general approval in the talks, but also gained numerous tips which have also indirectly left their mark on the structure of the handbook. The editors would at this point like to thank those in particular whose suggestions, opinions or criticisms were indeed noted but are not explicitly acknowledged, such as Martin Kay, Olga Kulagina, Kathleen McKeown, Igor Melčuk, Roger C. Schank, Wolfgang Wahlster, Donald Walker, Yorick Wilks and many others whom we approached with a view to collaboration, but who nevertheless felt they were not in a position to accept the offer. Similarly the editors request the understanding of those whose suggestions it was necessary to leave aside.

The assignation of tasks to a handbook — as is explained more fully below — leads to the fact that its lay-out mirrors a possible structuring of the subject as a whole. In view of the lack of specialist categorization in the field of Computational Linguistics, the editors were confronted with the task of outlining not only the structure of a volume, but also of a whole subject.

Establishing a structure was impeded by two factors: firstly by the absence of concentrated preliminary work; and secondly by constantly rapid development in the research area. So on the one hand it was necessary to find a comprehensive, factually appropriate and firm structure which contains the results of research and development attained so far in Computational Linguistics; on the other hand this structure must adopt a form so flexible that it allows for the classification of ongoing research. Problem areas were not to be omitted, neither were newly-found areas of research to be ignored.

In an effort to do justice to both requirements, the editors drew-up a systematic structure for the volume — which will be explained in further detail — along with thematic guidelines for the individual articles, which established the framework. It was left to the specialists invited to help work on the volume to provide the details of the framework.

The recruitment of contributors began in 1983. Leading representatives of the field were to present the various branches against the background of the current stage in their development. In order to keep subjective misjudgments to a minimum, particularly where the evaluation of individual problem areas is concerned, the editors aimed at a wide international co-operation so that the whole field of Computational Linguistics was covered rather than obtaining regional representation.

While the allocation and form of individual articles have seen inevitable compromises, it will be maintained in retrospect that it was necessary to only slightly alter the original structure intended for the subject, although, in view of the rapid advances made in Computational Linguistics, a revision of the structure would not have been surprising.

The handbook's treatment of the particularly concentrated development of *language-oriented Artificial Intelligence* in recent years is all too inadequate. It is in neither the interest of language-oriented Artificial Intelligence research nor of Computational Linguistics to strive for a clear dividing line between the two. On the contrary, the two fields must in future be considered in closer connection with one another.

The editors regret the delay in the publication of the handbook. The original intention was to publish as early as 1986, (COLING '86 in Bonn), then in 1988 (COLING '88 in Budapest). Concern for the quality, and efforts to ensure the whole field was covered led to a tense wait for reworks or for surveys which were regarded as imperative for the draft framework of the handbook. The editors were on no account prepared to compromise on the quality of the contributions. The relatively long editing time demands a retard between the individual articles which the user of the handbook notes in particular in the premature break from the quoted literature.

### 3. The Purpose of the Handbook

The handbook is addressed to everyone dealing with the use of data processing in languages as defined in linguistics and communication science (see the preface by Ungeheuer/Wiegand in Vol. 1.1 of this series of Handbooks).

Handbooks more or less automatically claim to be complete, and to present relevant or at least still valid knowledge. Alone the latent claims of this genre mean great demands on the skill of both the editors and the collaborators of handbooks which of course also hold for this handbook.

In detail, the purpose of the handbook is

— to describe the current level of international research in comprehensive surveys,



- to discuss various conceptional attempts to locate the subject within science in general, to determine its structure, and examine the way it relates to neighboring disciplines,
- to present a survey of literature on this subject, which is often difficult to obtain.
- to report on the use of Computational Linguistics in fields other than that of basic research, and
- to demonstrate perspectives for the further development of the subject.

These general requirements that ought to be met by a handbook are supplemented by two further conditions in the case of Computational Linguistics:

- (1) Computational Linguistics is a relatively young research area, and its autonomy is not yet fully indisputable. The principle question arises whether Computational Linguistics really is necessary as a distinct subject. Couldn't automatic lemmatizing, automatic syntactic analysis, etc. be included in appropriate chapters (in morphology, syntax, etc.) of language description, i. e., in linguistics? Couldn't an additional division of applied linguistics or information science be established for machine translation or automatic indexing? What justifies regarding Computational Linguistics as a distinct scientific research field?
- (2) So far, the field has not been structured; this presents a problem yet to be solved. There have been no comprehensive attempts to integrate the various aspects of Computational Linguistics that could act as guidelines. Attempts at the structure have so far been one-sided, remained restricted to specific areas or were out-dated before they were successful (Hays 1967; Ungeheuers' Man-Machine-Model, Ungeheuer 1971; Kay in Sparck Jones/Kay 1973; Bátori 1977; Lenders 1980; Tennant 1981; Bátori/Krause/Lutz 1982; Grishman 1986). Nonetheless, these attempts do contain several good approaches that have also been taken up in this volume, compared with each other, and extended.

Under these circumstances, it is quite feasible, perhaps even probable, that the handbook does not cover all possible interests and topics; however, the editors are aiming at presenting a complete survey of the whole field that would at least enable the classification and further studying of specific problems.

The handbook should not be regarded as a textbook introducing the student to the various established methods of, say, machine text processing; however, it contains references to training colleges, and, in the bibliography, a list of publications that would be of use in further studying of specific topics.

Finally, it has to be pointed out that this handbook — unlike, for instance, the first two volumes of the handbook series — is not related to a specific language, like German. Instead, as in the case of the volumes *Semiotics*, *Philosophy of language*, *Lexicography*, etc., the attempt is made to ensure a presentation of the systems and processing methods dealt with, which is independent of individual languages.

## 4. On the Legitimation of Computational Linguistics

Even if Computational Linguistics is not regarded as a subdiscipline, but, less ambitiously, as a research field, the question remains: what justifies the delimitation of Computational Linguistics within (or separately from) linguistics?

The editors see the justification for Computational Linguistics becoming a distinct scientific field above all from a pragmatic point of view: the enormous increase in research and publication activities. Due to the lack of appropriate publication organs

and the splitting up of the field, many research results have only been publicized in the immediate vicinity of the respective linguistic research area. The implications of these results were often not appreciated, and above all they were not able to spread beyond traditional scientific fields. The results of the field are scattered about, and they are not in any way compiled. The practical objective of the handbook is to bridge this gap. But even in sectors that enjoy a better infrastructure the extent of specialized literature that has to be dealt with has increased so rapidly that the orientation and classification of one's own work becomes difficult even in the various subdivisions of Computational Linguistics, like syntax parsing or production of concordances, and cannot be left to the individual philological zeal of a scientist without endangering the innovative handing down of science. In order to guarantee the continuity of scientific research, it is necessary to record the newly created computer-orientated linguistic research comprehensively, thus offering guidelines to individual researchers, tutors, and students of the field.

However, this pragmatic justification of Computational Linguistics remains empty as long as one does not succeed in making plausible the coherence of the field. Recognizing the pragmatic necessity does not become superfluous, but merely shifts the quest for a foundation of Computational Linguistics guided by theoretical principles.

Justifying Computational Linguistics as a distinct research field ultimately implies examining its inner relations. In what way are the various research branches connected that are commonly (and in this handbook) combined in Computational Linguistics?

A theoretically orientated definition of the location of Computational Linguistics as related to linguistics would start out from the important role of data processing in the research field and specify the interdisciplinary links between language and the computer from a linguistic angle. The autonomy of Computational Linguistics can be justified by innovations that are assigned to four levels of knowledge:

- (1) The *new methods*: Computational Linguistics came into being when the new techniques of electronic data processing were also used for dealing with old problems which were known from the separate philologies. These methods are indisputably new and different from the usual philological methods. This is demonstrated in a typical way by the development of the field of lexicography.
- (2) *New insights*: Examining old problems with a new research instrument enables new, hitherto inconceivable insights. Precision and the amount of data that can be dealt with by machine processing open up new perspectives and enable qualitatively new statements. An example for this is computer-aided authorship determination.
- (3) The advent of computer technology has led to several *new types of problems*. Apart from direct computer application, like machine translation or automatic indexing, secondary processes that have already arisen or are arising from the use or partial use of computers in research and can be dealt with by computers, like lemmatizing, must also be mentioned.
- (4) Finally, a *new structuring of the field* has evolved. New methods, insights, and types of problems lead to new affinities that cannot be included in conventional linguistic field structuring, for instance the process problem in man-machine communication. For an adequate description of the field, a complete new field structuring will become necessary (based on the fundamental distinction between *description* and *simulation*), and it would clearly describe Computational Linguistics as a field separate from conventional linguistics.

Of course these four levels do not have equal degrees of importance; for instance, methodical innovations alone would certainly not be sufficient, but their argumentative strength increases in the given context.

It is often difficult to access in a detailed way how the new (exact) computer-aided research methods have enabled new insights and new problem solutions, or to what extent types of problems and the restructuring of the field are related to each other. However, one can observe that the higher levels of knowledge are to a greater degree responsible for the constituting of the new linguistic subdiscipline of Computational Linguistics than the lower ones.

The fourth level of abstraction, the new structuring of the field, becomes decisive inasmuch as it systematically accumulates the innovations of the lower levels and brings together and relates hitherto isolated phenomena and problems lying far apart. Structuring is not only an inventory but also supplies an explication of the relations and common features within the field; it opens up new perspectives and gives the field the quality of autonomy. In this sense, just how firmly Computational Linguistics is established depends on its structuring.

The editors believe that the quality and the number of innovations these four levels comprise justify the autonomy of Computational Linguistics with respect to linguistics in a sufficient way. Nevertheless, in view of the linguistic dominance as far as defining the types of problems is concerned, Computational Linguistics remains closely linked with linguistics.

This orientation towards linguistics must be emphasized especially because the inherent interdisciplinary research field can at least partly be defined and structured in the context of other research directions. Two of these alternative ways of structuring Computational Linguistics that have been discussed intensively recently, cognitive science and artificial intelligence, ought to be briefly described.

*Cognitive science* denotes a field of knowledge in which the more sophisticated human intellectual abilities are examined, including the use of language. The field is especially complicated within the sector of language use and language behavior, and, quite appropriately, several traditional and new (sub-)disciplines lay claim to its research. Cognitive science combines all those disciplines that can contribute to research on more elaborate cognitive abilities since, as far as solving problems is concerned, it is not important who has contributed what to the solution or which discipline is responsible for the research work done.

So cognitive science would bring together in an organized way research that has hitherto been hindered by disciplinary boundaries, and it would create new, stimulating connections without having to deal with cumbersome delimitation and competence issues.

The promising openness of the concept of an (as yet hardly institutionalized) cognitive science is, however, also accompanied by the necessity of an inner structuring. For, on the one hand, it ought to be appreciated that image processing and language learning are regarded as immediate neighbors, but on the other hand, the topic of language behavior must be dealt with as a unit, so that in the end the autonomy of Computational Linguistics is not affected by cognitive science.

So cognitive science and Computational Linguistics are not rivals or mutually exclusive conceptions. The accentuated interdisciplinarity, especially the extending of the horizon towards psychology and the study of volitional aspects of language (ranging, beyond pure information conveying) are also attractive perspectives for Computational Linguistics, especially in the area of language simulation.

*Artificial intelligence* (= AI) deals with problem solving, and the phenomenon of language is an inevitable result of the research methods applied: problems must be formulated in a straightforward way before their solution can be attempted, both in heuristic searching and automatic theorem proving. So language use is a central manifestation of intelligent behavior. Artificial intelligence is primarily orientated towards computer science, and it is inspired directly by the concepts of computer science in synthesizing the problem field (Barr and Feigenbaum 1981, 3).

Whereas cognitive science copes, or hopes to cope, with the problem of surmounting disciplinary boundaries by organizing all disciplines that deal with cognition problems — perhaps tentatively — in a tightly woven communication network, thus perforating disciplinary boundaries, artificial intelligence eliminates them totally by laying claim to relevant knowledge areas itself and managing with the problem of interdisciplinarity within information science, or at least it hopes to do so.

Since Computational Linguistics and artificial intelligence are devoted to separate directions of research, this leads to characteristic differences which constitute distinct aspects of research:

- (1) Artificial intelligence does not grant a special status to language as one of the human capacities. It is a basic tenet of linguistics (and ipso facto of Computational Linguistics too) that human language is irrevocably allied with thought and is therefore qualitatively different from many other (non-verbal) displays of human intelligence. From a linguistic point of view language difficulties justify scientific research as an end in itself. AI views the discussion of language problems as being justifiable only within a wider context of problem solving (cf. Chomsky, *Reflections on Language* 1976, 22).
- (2) Artificial Intelligence requires neither an autonomous level of language analysis nor a Competence-Performance Dichotomy as is usually the case in linguistics. There the directly available formulation of language is confronted by an abstract representation of knowledge (cf. Wahlster 1982).

In the field of language behaviour models there are fundamental overlaps between Artificial Intelligence and Computational Linguistics: their relationship is therefore to be regarded as one of vicinal areas with overlaps, in which it is not so much the mutual division which is to be encouraged, but rather the reciprocal exchange of experience and knowledge. (The relationship of Artificial Intelligence and Computational Linguistics is assessed differently: some representatives of Artificial Intelligence are quite prepared to adopt recognitions of linguistic interest as their own (cf. Habel 1982).

## 5. The Structuring of the Handbook

### 5.1 Structuring Principles

In a similar way to justifying the autonomy of Computational Linguistics, in structuring the field, one takes as a basis the prominent role of data processing in linguistic research.

This concept of the role of electronic data processing is expressed in the central chapters of the handbook in the clear distinction between two aspects that have been extracted from various conceptional designs and the practice of Computational Linguistics. For one thing, the computer is regarded as an aid in *describing* language phenomena which can be employed to prepare the empirical foundations of theories relating to languages, especially linguistic theories. Description means both inductive processes which start out from a given language phenomenon and detect specific regularities in it, and deductive processes applied in examining language phenomena, incorporating a verification of assumed regularities by tests.

On the other hand, the computer is not only seen as an aid in describing language, but as a language user itself, i. e., as a unit that can react to the perception of language phenomena and can also produce these phenomena (*computer simulation*). In this respect, the data processing equipment also acts as an instrument for developing theories, theories on language behavior.

So language description and language simulation are the two structuring aspects of Computational Linguistics, which are linked in various ways. This division can also be observed in the different areas of application.

The system presented here does not claim to be final, but in the opinion of the editors it reflects the latest developments in the field. The editors hope that the structuring suggested in this volume will prove to be useful and extendable beyond the scope of the handbook.

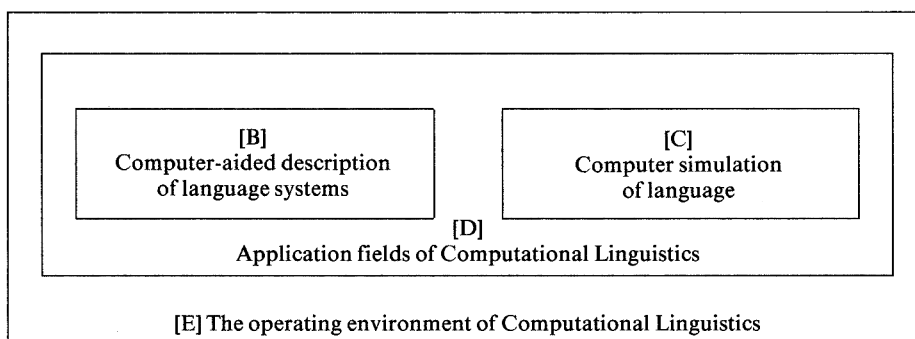
## 5.2 The Structure of the Handbook in Detail

The two central topics of Computational Linguistics, description and simulation, are presented in chapters IV—VIII of the handbook. They are preceded by some theoretical-methodical chapters I—III, in which the development of the field is described, and they are followed by three chapters (IX—XI) which include the areas of application of Computational Linguistics. Chapter XII on working practice concludes the handbook.

The handbook thus consists of the following five main parts:

- [A] The development, theory, and methods of Computational Linguistics (ch. I—III)
- [B] Computer-aided description of language systems (ch. IV—VI)
- [C] Computer simulation of language (ch. VII—VIII)
- [D] Application fields of Computational Linguistics (ch. IX—XI) and
- [E] Operating environment of Computational Linguistics (ch. XII).

It has already been mentioned that in the main part [A] there is an attempt to place the collective study of Computational Linguistics in a wider, historical and epistemological, external framework. The internal classification of the individual main parts is as follows



The varied levels of information in the individual main parts becomes particularly clear when we consider for how long the information gained may be of profitable use. While linguistic description [B] and simulation [C] present relatively constant information, the information in the area of application ages more rapidly. Even more dramatic is the change in the information in the practical working field [E].

The handbook presents a bibliography of some 3 700 titles on Computational Linguistics, compiled from the references of the individual articles. The compilation of the bibliography was done by Friedrich Wetter, to whom the editors want to express their warmest thanks. Although the editors cannot demand complete entirety, the handbook they present contains the largest comprehensive, systematic collection of literature on Computational Linguistics.

## 6. Computational Linguistics: Dream and Reality

It would certainly be inappropriate to claim that Computational Linguistics is in a position to provide a complete description of the processes of understanding language, which may also serve as a basis for the application of computer programs in natural language. But it has become clear that Computational Linguistics can, and indeed to some extent does, provide a basis, without which it will not be demonstrated that the range of basic research in Computational Linguistics is also understood beyond the narrow limits of its field.

With this in mind we must acknowledge the visionary analysis of Gerold Ungeheuer who introduced the new discipline at a time when the connection between language and the computer was still obscured by misconceptions and prejudices. In two respects he was ahead of his time:

- (1) At a time — and this time is by no means very long ago — when academics ignored or rejected it, he recognized and propagated the computer as a working instrument for language research.
- (2) At a time — and this time too is by no means very long ago — when scientists still regarded the processing of human information as a precedent to the general processing of symbols, he recognized that machine language processing can only be achieved within a larger communicative framework.

This handbook indicates how the world has undergone radical change since this time, and how appropriate Ungeheuer's analysis was in 1971.

March, 1989

István S. Bátori (Koblenz)  
Winfried Lenders (Bonn)  
Wolfgang Putschke (Marburg)

## Vorwort

### 1. Zur Bezeichnung des Fachgebietes

In den vergangenen vier Jahrzehnten hat sich aus kleinsten Anfängen das Fachgebiet entwickelt, für das im vorliegenden Handbuch die Bezeichnung *Computerlinguistik* bzw. *Computational Linguistics* (CL) gewählt worden ist, das aber in der Praxis von Forschung, Lehre und Anwendung mit verschiedenen Namen wie *Linguistische Datenverarbeitung*, *Computative Linguistik*, *Sprachdatenverarbeitung*, *Linguistische Informatik*, *Informationslinguistik*, *Literary and Linguistic Computing*, *Language Data Processing* usw. benannt wird und zu dem auch die sprachbezogenen Teile der Künstlichen-Intelligenz-Forschung gehören. In den verschiedenen Namen des Faches spiegeln sich die Interessen so unterschiedlicher Disziplinen wie Linguistik, Informationswissenschaft, Informatik, Psychologie usw., sowie die der verschiedenen Philologien und einiger Anwendungsfelder wie Sprachübersetzung, Informationssysteme, philologische Textverarbeitung u. a. wider. Bei aller Verschiedenheit der Bezeichnung hat dieses Fach jedoch einen einheitlichen Gegenstand: Sprache und sprachliche Phänomene, bei deren Erforschung als besonderes Forschungsinstrument der Computer eingesetzt wird. Mit dem Namen des vorliegenden Handbuches wurde angesichts dieser Vielfalt von Aspekten diejenige Bezeichnung ausgewählt, die den Herausgebern am geeignetsten erschien, um unterschiedlich akzentuierte Arbeitsgebiete zusammenzuführen. Zudem ist die Bezeichnung *Computational Linguistics* als Name der größten internationalen Fachvereinigung eingeführt und daher international verständlich. Im deutschen Kontext wird zunehmend die Bildung *Computerlinguistik* verwendet, da diese nicht wie andere Bezeichnungen nur regional gebräuchlich ist.

### 2. Über das Entstehen dieses Handbuches

Es war der früh verstorbene Mitherausgeber der Reihe *Handbücher zur Sprach- und Kommunikationswissenschaft*, Gerold Ungeheuer, der die Notwendigkeit dieses Handbuches als erster erkannt und sich für seine vorrangige Aufnahme in die Reihe eingesetzt hat. Seinem Andenken widmen die Herausgeber diesen vorliegenden Band.

Die ersten Orientierungsgespräche fanden 1981 statt. Ein Leitfaden für das Handbuch und ein Entwurf für seine Gliederung wurden erarbeitet, auf den Computerlinguistik-Kongressen in Prag (1982) und Palo Alto/Stanford (1984) verteilt und mit Kollegen diskutiert. In den Gesprächen haben die Herausgeber nicht nur eine breite Zustimmung registriert, sondern auch zahlreiche Ratschläge bekommen, die indirekt auch in die Gliederung des Handbuches eingeflossen sind. Die Herausgeber möchten an dieser Stelle insbesondere denen danken, deren Vorschläge, Stellungnahmen oder Kritik zwar beherzigt aber ohne explizite Antwort geblieben sind, wie Martin Kay, Olga Kulagina, Kathleen McKeown, Igor Melčuk, Roger C. Schank, Wolfgang Wahlster, Donald Walker, Yorick Wilks und vielen anderen, die wir wegen Mitwirkung angesprochen haben, die sich jedoch nicht in der Lage sahen, der Einladung zu folgen. Gleichzeitig möchten die Herausgeber bei allen Kolleginnen und Kollegen um Verständnis werben, deren Anregungen zurückgestellt werden mußten.

Aus der Aufgabenstellung eines Handbuchs folgt, wie unten noch näher erörtert wird, daß seine Gliederung unausweichlich einer möglichen Gliederung des gesamten Faches gleichkommt. Angesichts einer noch fehlenden fachlichen Systematik der Computerlinguistik standen die Herausgeber daher vor der Aufgabe, nicht nur die Gliederung für einen Band, sondern gleichzeitig eine Gliederung für ein ganzes Fach zu entwerfen.

Die Aufstellung einer Gliederung war durch zwei Umstände erschwert, erstens durch das Fehlen von konsensfähigen Vorarbeiten und zweitens durch die anhaltende rasche Entwicklung des Forschungsfeldes. Einerseits mußten also eine umfassende, sachlich angemessene und feste Gliederung gefunden werden, die die bisherigen Ergebnisse der Forschungs- und Entwicklungsarbeit in der Computerlinguistik abdeckt, andererseits mußte diese Gliederung so flexibel gestaltet werden, daß sie auch die Einordnung der laufenden Forschung erlaubt. Es durften weder Problemdomänen ausgespart, noch neu entstandene Forschungsrichtungen ignoriert werden.

Um beiden Forderungen gerecht zu werden, erstellten die Herausgeber eine systematische Gliederung für den Band, die noch näher erörtert wird, sowie thematische Richtlinien für die einzelnen Artikel, die den Rahmen festlegten. Den zur Mitarbeit eingeladenen Spezialisten wurde es überlassen, diesen Rahmen im einzelnen aufzufüllen.

Die Anwerbung der Mitarbeiter begann 1983. Führende Fachvertreter sollten die einzelnen Teilgebiete auf dem aktuellen Entwicklungsstand darstellen. Um subjektive Fehleinschätzungen vor allem in der Gewichtung einzelner Problemdomänen zu minimieren, strebten die Herausgeber eine breite, internationale Kooperation an, wobei regionale Repräsentativität der konsequenten Abdeckung der gesamten Computerlinguistik nachgeordnet wurde.

Obwohl Kompromisse bei der Vergabe und Gestaltung der einzelnen Artikel unvermeidlich gewesen sind, ist im Rückblick festzustellen, daß die ursprüngliche Gliederung des Bandes und der ursprünglich vorgesehene thematische Rahmen der einzelnen Artikel nur wenig geändert werden mußten, obwohl, angesichts des stürmischen Vormarsches der Computerlinguistik, eine Revision der Gliederung nicht verwunderlich gewesen wäre.

So reflektiert das Handbuch die besonders intensive Entwicklung der *sprachorientierten Künstlichen Intelligenz* der letzten Jahre nur unzulänglich. Es ist weder im Interesse der sprachorientierten Künstlichen Intelligenz, noch der Computerlinguistik, hier eine scharfe Abgrenzung anzustreben. Im Gegenteil, beide Bereiche müssen künftig stärker aufeinander bezogen werden.

Die Herausgeber bedauern die Verzögerungen bei der Publikation des Handbuchs. Ursprünglich strebte man die Veröffentlichung bereits für 1986 (COLING '86 in Bonn), dann 1988 (COLING '88 in Budapest) an. Die Sorge um die Qualität und das Streben nach Abdeckung des gesamten Gebietes führten zum angespannten Warten auf Überarbeitungen oder auf Übersichtsartikel, die für den konzeptuellen Rahmen des Handbuchs als unerläßlich angesehen wurden. In keinem Fall waren die Herausgeber bereit, Konzessionen bezüglich der Qualität der Beiträge in Kauf zu nehmen. Die relativ lange redaktionelle Bearbeitungsdauer bedingt, daß eine zeitliche Diskrepanz zwischen den einzelnen Artikeln vorhanden ist, die sich für den Benutzer des Handbuchs insbesondere in dem frühen Abbruch der zitierten Literatur manifestiert.

### 3. Aufgaben des Handbuchs

Das Handbuch richtet sich an alle, die sich mit Problemen der Anwendung der Datenverarbeitung auf Sprache im Sinne der Sprach- und Kommunikationswissenschaft befassen (vgl. das **Geleitwort von Ungeheuer/Wiegand im Bd. 1.1 dieser Reihe**).



Handbücher erheben Anspruch auf Vollständigkeit und Aktualität. Bereits aus diesen latenten Ansprüchen der Gattung ergeben sich im allgemeinen hohe Anforderungen sowohl an Mitarbeiter als auch an Herausgeber von Handbüchern, die selbstverständlich auch für dieses Handbuch gelten. Im einzelnen besteht die Aufgabe dieses Handbuches darin,

- den gegenwärtigen Stand der Forschung in übersichtlicher Form darzustellen,
- verschiedene konzeptionelle Entwürfe zum Standort des Faches in der Gesamtheit der Wissenschaften, zu seiner Struktur und solche zu seinen Beziehungen zu Nachbardisziplinen zu erörtern,
- die teils schwer zugängliche Literatur des Faches zu erschließen,
- über vorhandene Anwendungen außerhalb der wissenschaftlichen Grundlagenforschung zu berichten und
- die Perspektiven für die weitere Entwicklung des Faches aufzuzeigen.

Zu den allgemeinen Forderungen an ein Handbuch kommen im Falle des Handbuchs *Computerlinguistik* zwei weitere Aspekte, die zu berücksichtigen sind.

- (1) *Computerlinguistik* ist ein junges Forschungsgebiet, und ihre Eigenständigkeit ist noch nicht selbstverständlich. Man kann die Frage grundsätzlich aufwerfen: Braucht man *Computerlinguistik* überhaupt? Könnte nicht automatische Lemmatisierung, maschinelle Syntexanalyse usw. in den jeweils passenden Kapiteln (in der Morphologie, in der Syntax usw.) der Sprachbeschreibung, also innerhalb der Linguistik, aufgenommen werden? Könnte man nicht für Maschinelles Übersetzen oder Automatisches Indexing eine zusätzliche Abteilung in den sog. *Angewandten Sprachwissenschaften* bzw. der *Informationswissenschaft* vorsehen? Was berechtigt überhaupt, *Computerlinguistik* alseigenständiges Forschungsfeld anzusehen?
- (2) Die Gliederung des Feldes ist noch nicht gegeben, sondern sie sollte erst gefunden werden. Es gibt keine Vorbilder, keine wirklich umfassenden Integrationsversuche in der Computerlinguistik. Die bisherigen Gliederungsversuche waren einseitig, blieben auf Teilbereiche beschränkt oder veralteten, bevor sie sich durchsetzten (Hays 1967; Ungeheuers Mensch-Maschinen-Modell, Ungeheuer 1971; die Ansätze von Kay in Sparck Jones/Kay 1973; Bátori 1977; Lenders 1980; Tenant 1981; Bátori/Krause/Lutz 1982; Grishman 1986.) Nichtsdestoweniger enthalten diese Versuche viele gute Gliederungsansätze, die auch in diesem Band aufgegriffen, untereinander abgeglichen und weiterentwickelt wurden.

Unter diesen Umständen ist es durchaus denkbar, ja wahrscheinlich, daß das Handbuch nicht alle möglichen Interessen und Themen berücksichtigt; die Herausgeber streben jedoch an, erstmalig einen vollständigen Überblick über das gesamte Wissensgebiet anzubieten, der mindestens die Einordnung und weitere Verfolgung der einzelnen Fragen ermöglicht.

Das Handbuch ist nicht als Lehrbuch zu verstehen, das den Lernenden in die verschiedenen etablierten Techniken etwa der maschinellen Textverarbeitung einführt; es enthält jedoch weiterführende Hinweise auf Lehrstätten und, in der Bibliographie Hinweise auf Publikationen, über die ggf. die einzelnen Bereiche vertieft werden können.

Schließlich ist zu bemerken, daß sich dieses Handbuch — anders als beispielsweise die beiden ersten Bände dieser Handbuchreihe — nicht auf eine Sprache, etwa das Deutsche, bezieht. Vielmehr wird versucht, — ähnlich wie es in den Bänden *Semiotik*, *Sprachphilosophie*, *Lexikographie* der Fall ist oder der Fall sein wird — eine von Einzelsprachen unabhängige Darstellung der behandelten Systeme und Veranstaltungsmethoden zu gewährleisten.

## 4. Über die Legitimation der Computerlinguistik

Aber selbst wenn die CL nicht als Teildisziplin, sondern etwas weniger anspruchsvoll als Forschungsfeld begriffen wird, bleibt die Frage offen: Was berechtigt die Abgrenzung der Computerlinguistik innerhalb (oder neben) der Linguistik?

Die Herausgeber sehen die Berechtigung für die Verselbständigung der Computerlinguistik in erster Linie im pragmatischen Bereich, in den enorm anwachsenden Forschungs- und Publikationsaktivitäten. In Ermangelung von zuständigen Publikationsorganen und infolge der Zersplitterung des Feldes sind nämlich viele Forschungsergebnisse häufig in partikularlinguistischer Umgebung publiziert worden. Diese Ergebnisse konnten dort vielfach in ihrer gesamten Tragweite nicht erfaßt werden, und vor allem vermochten sie oft nicht, die Grenzen der traditionellen Fachgebiete zu überspringen. Die Ergebnisse des Feldes sind zerstreut und nirgends kompiliert. Diese Lücke zu füllen, ist der praktische Anlaß des Handbuchs. Aber selbst in Teilbereichen mit besserer Infrastruktur hat der Umfang der zu bewältigenden Fachliteratur derart zugenommen, daß die Orientierung und die Einordnung der eigenen Arbeit selbst in den einzelnen Teilbereichen (der CL), wie Syntaxparsing oder Konkordanzerstellung, schwerfällt und ohne Gefährdung der innovationsstiftenden Wissenstradierung nicht mehr der individuellen philologischen Beflissenheit der einzelnen Forscher überlassen werden kann. Im Interesse der Kontinuität der wissenschaftlichen Forschung ist es notwendig, die neu entstandene computerorientierte linguistische Forschung ganzheitlich zu erfassen und hierdurch den einzelnen Forschern, Lehrern und Studierenden des Feldes Orientierungshilfe anzubieten.

Diese pragmatische Legitimation der Computerlinguistik bleibt jedoch solange leer, als es nicht gelingt, die immanente Kohärenz des Feldes plausibel zu machen. Die Erkennung der pragmatischen Notwendigkeit erübrigt sich nicht, sondern sie macht die Suche nach einer von theoretischen Prinzipien geleiteten Fundierung der CL noch dringender.

Die Frage nach der Berechtigung der Computerlinguistik ist letztendlich die Frage nach der Verbindung ihrer inneren Zusammenhänge: Was verbindet die einzelnen Forschungszweige, die üblicherweise (und auch in diesem Handbuch) als Computerlinguistik zusammengefaßt werden?

Auch eine theoretisch orientierte Standortbestimmung der Computerlinguistik gegenüber der Linguistik wird von der prominenten Rolle der Datenverarbeitung in dem Forschungsfeld ausgehen und die interdisziplinären Verbindungen zwischen Sprache und Computer von der linguistischen Seite her präzisieren. Die Eigenständigkeit der Computerlinguistik läßt sich durch Neuerungen, die vier Erkenntnisebenen zugeordnet werden können, begründen:

- (1) Durch die *neuen Methoden*. CL begann, als die neuen Arbeitsmethoden der elektronischen Datenverarbeitung für die Bewältigung der alten von den Einzelphilologen her bekannten Probleme angewandt wurden. Diese Methoden sind unumstritten neu und anders als die tradierte philologische Arbeitsweise. Mittlerweile gibt es umfangreiche Erfahrungen mit diesen neuen Methoden. Prototypisch hierfür ist die Entwicklung im Bereich der *Lexikographie*.
- (2) Durch die *neuen Einsichten*. Die Untersuchung alter Probleme mit dem neuen Forschungsinstrument eröffnet neue, vorher nicht vorstellbare Einsichten. Die Präzision sowie die durch die maschinelle Verarbeitung erfaßbare Datenmenge eröffnen neue Perspektiven und ermöglichen neue Aussagequalitäten. Ein Beispiel hierfür ist die *Autorenschaftsbestimmung mit Computerhilfe*.
- (3) Es gibt eine Reihe von neuen *Aufgabenstellungen*, die erst durch die neue Computertechnologie aufgekomen sind. Hier ist nicht nur an die unmittelbare Compu-

teranwendung zu denken, wie Maschinelle Sprachübersetzung oder Automatisches Indexing, sondern auch an die sekundären Aufgabenstellungen, die bereits im Zusammenhang mit Computeranwendungen bzw. mit Computereinsatz in der Forschung entstanden sind und entstehen und ebenfalls mit Hilfe von Computern bewältigt werden, wie z. B. Automatische Lemmatisierung.

- (4) Schließlich ergibt sich eine *Neugliederung des Feldes*. Durch neue Methoden, Einsichten und Aufgabenstellungen entstehen neue Affinitäten, die mit Hilfe der traditionellen linguistischen Feldgliederungen nicht erfaßt werden können, wie z. B. die Prozeß-Problematik oder die Mensch-Maschine-Interaktion. Für eine adäquate Feldbeschreibung wird eine neue eigenständige Feldgliederung notwendig sein (mit der grundlegenden Unterscheidung zwischen *Deskription* und *Simulation*), die das neue Gebiet von der Linguistik klar unterscheidet.

Diese vier Ebenen sind von unterschiedlicher Bedeutung. So würden die methodischen Neuerungen allein für die Gründung der CL sicherlich nicht ausreichen; sie gewinnen jedoch in ihrem Zusammenhang an argumentativer Kraft.

In Einzelheiten ist es oft schwierig zu sagen, wie weit die neuen (exakten) computer-gestützten Forschungsmethoden die neuen Einsichten bzw. Problemlösungen bewirkt haben oder wie Aufgabenstellungen und Neugliederungen des Feldes zusammenhängen. Es läßt sich jedoch behaupten, daß für die Konstituierung der Computerlinguistik die höheren Erkenntnisebenen stärker verantwortlich sind als die niederen.

Entscheidend ist die letzte Abstraktionsebene, die neue Feldgliederung, die die Neuerungen der niederen Ebenen systematisch akkumuliert und die bisher isolierte, voneinander entfernte Erscheinungen oder Problembereiche zusammenführt und in Beziehung zueinander setzt. Die Gliederung ist nicht nur ein Inventar, sondern gleichzeitig auch eine Explikation der Zusammenhänge und Gemeinsamkeiten des Feldes, sie eröffnet neue Perspektiven und verschafft dem Feld die Qualität der Eigenständigkeit. Wie fest die Computerlinguistik fundiert ist, steht und fällt in diesem Sinne mit ihrer Gliederung.

Die Herausgeber meinen, daß die Qualität und die Menge der Neuerungen auf diesen vier Erkenntnisebenen die Eigenständigkeit der Computerlinguistik gegenüber der Linguistik hinreichend begründen. Nichtsdestoweniger bleibt durch die linguistische Dominanz der Fragestellungen die Computerlinguistik mit der Linguistik eng verbunden.

Diese linguistische Zuordnungsrichtung ist deshalb besonders hervorzuheben, weil das inhärent interdisziplinäre Forschungsfeld zumindest partiell auch mit anderen Erkenntnisinteressen abgegrenzt und zugeordnet werden kann. Zu zwei von diesen alternativen Zuordnungsrichtungen soll hier kurz Stellung genommen werden, die in der letzten Zeit besonders intensiv diskutiert werden, nämlich zur Cognitive Science und zur Künstlichen Intelligenz.

Unter dem Namen *Cognitive Science* wird ein Wissensbereich verstanden, in dem die höheren geistigen Fähigkeiten des Menschen, darunter auch die Sprachanwendung, erforscht werden. Das Feld ist insbesondere im Bereich der Sprachanwendung und des Sprachverhaltens so komplex, daß sich für seine Erforschung mit Recht eine Reihe von (traditionellen und neuen) (Teil-)Disziplinen berufen fühlen (Bobrow/Collins 1977). Cognitive Science führt alle Disziplinen zusammen, die bei der Erforschung höherer geistiger Fähigkeiten einen Beitrag leisten können, denn für die Problemlösung an sich ist es irrelevant, wer wieviel zur Lösung beitrug bzw. wie die für die gefundene Lösung verantwortliche Disziplin heißt.

Cognitive Science würde also die bisher durch Disziplinengrenzen gehemmte Forschung organisatorisch zusammenbringen und neue, anregende Nachbarschaften schaffen, ohne sich um Abgrenzungs- und Kompetenzfragen kümmern zu müssen.

Die begrüßenswerte Offenheit des Konzepts der Cognitive Science wird jedoch auch von der Notwendigkeit der inneren Strukturierung begleitet. Denn einerseits ist es positiv, daß Bildverarbeitung oder Sprachlernen als unmittelbare Nachbarn angesehen werden, andererseits aber gehört die Sprachverhaltensproblematik doch enger zusammen, so daß man letztlich die Eigenständigkeit der Computerlinguistik durch Cognitive Science nicht berührt.

Cognitive Science und Computerlinguistik sind daher keine Rivalen oder einander wechselseitig ausschließende Konzepte. Die akzentuierte Interdisziplinarität, insbesondere die Horizonterweiterung in Richtung der Psychologie und die Betrachtung der volitionalen Aspekte der Sprache (über die reine Informationsvermittlung hinaus) sind auch für die Computerlinguistik (insbesondere im Sprachsimulationsbereich) attraktive Perspektiven.

*Künstliche Intelligenz* (= KI) widmet sich der Problemlösung und kommt dabei vor allem durch methodischen Zwang zu dem Phänomen Sprache: Probleme müssen nämlich unausweichlich formuliert werden, bevor ihre Lösung vorgenommen wird, gleichgültig, ob es sich um heuristisches Suchen oder um automatische Theorembeweise handelt. Daher ist die Verwendung eine zentrale Manifestation von intelligentem Verhalten. Die Zuwendungsrichtung der KI ist primär die Informatik, denn sie versteht sich als einer ihrer Bereiche (Barr, Feigenbaum 1981, 3), und sie wird bei der Synthese des Problemfeldes direkt durch die Konzepte der Informatik inspiriert.

Während Cognitive Science das Problem der Überwindung der Disziplingrenzen dadurch bewältigt (bzw. zu bewältigen hofft), daß sie alle an der Kognitionsproblematik arbeitenden Disziplinen, u. U. auch auf Verdacht, nebeneinander in einem engen Kommunikationsnetz verbunden ansiedelt, also die Disziplingrenzen durchlässig macht, eliminiert die Künstliche Intelligenz sie gänzlich, indem sie die relevanten Wissensbereiche für sich hereinholt und die Problematik der Interdisziplinarität innerhalb der Informatik bewältigt (bzw. zu bewältigen hofft).

Durch die unterschiedlichen Zuwendungsrichtungen der CL und der KI ergeben sich charakteristische Unterschiede, die als verschiedene Forschungsaspekte verstanden werden:

- (1) KI räumt der Sprache unter den menschlichen Fähigkeiten keinen Sonderstatus ein. In der Linguistik (und eo ipso auch in der Computerlinguistik) geht man hingegen davon aus, daß die Sprache des Menschen untrennbar mit dem Denken verbunden und daher qualitativ anders ist als sämtliche anderen (nicht verbalen) menschlichen Intelligenzleistungen. Sprachliche Probleme legitimieren ihre wissenschaftliche Erforschung aus der linguistischen Perspektive gesehen bereits von sich aus. Aus der Sicht der KI ergibt sich die Legitimation für die Auseinandersetzung mit der Sprachproblematik erst in einem größeren Problemlösungszusammenhang (vgl. Chomsky, *Reflections on Language* 1976, S. 22).
- (2) KI benötigt weder eine autonome sprachliche Analyseebene noch die in der Linguistik übliche Kompetenz-Performanz-Dichotomie: Hier steht die unmittelbar zugängliche sprachliche Formulierung einer abstrakten Wissenspräsentation gegenüber (vgl. Wahlster 1982).

Im Bereich der Sprachverhaltensmodelle gibt es zwischen Künstlicher Intelligenz und Computerlinguistik in der Tat wesentliche Überlappungen: Ihr Verhältnis ist daher als eine Nachbarschaft mit Überdeckungen aufzufassen, wobei nicht so sehr die gegenseitige Abgrenzung, sondern im Gegenteil der wechselseitige Austausch von Erfahrungen und Einsichten zu fördern sind. (Das Verhältnis KI zu CL wird unterschiedlich beurteilt; manche KI-Vertreter sind durchaus bereit, sich auch linguistische Erkenntnisinteressen zu eigen zu machen [vgl. Habel 1982].)

## 5. Gliederung des Handbuches

### 5.1 Prinzipien der Gliederung

Wie bei dem Versuch, die Eigenständigkeit der Computerlinguistik zu begründen, geht man auch bei der Gliederung des Feldes von der prominenten Rolle der Datenverarbeitung für die linguistische Forschung aus:

Diese Vorstellung von der Rolle der elektronischen Datenverarbeitung kommt in den zentralen Kapiteln dieses Handbuches in der grundlegenden Unterscheidung von zwei Aspekten zum Ausdruck, die sich als Extrakt aus verschiedenen konzeptionellen Entwürfen und aus der Praxis der Computerlinguistik ergeben. Zum einen wird der Computer als Hilfsmittel bei der *Beschreibung* sprachlicher Phänomene verstanden, mit welchem die empirische Grundlage sprachbezogener Theorien, besonders linguistischer Theorien, aufbereitet werden kann. Mit *Beschreibung* ist sowohl das induktive Vorgehen gemeint, bei welchem von vorgefundenen Sprachphänomenen ausgegangen und die darin enthaltenen Regularitäten ermittelt werden sollen, als auch das deduktive Verfahren bei der Untersuchung sprachlicher Phänomene, wobei angenommene Regeln durch Tests verifiziert werden.

Zum anderen wird der Computer nicht als Hilfsmittel zur Sprachbeschreibung, sondern selbst als Sprachbenutzer aufgefaßt, d. h. als eine Größe, die auf die „Wahrnehmung“ sprachlicher Erscheinungen reagieren und die sprachliche Erscheinungen hervorbringen kann (*Computersimulation*). Auch in dieser Hinsicht dient die Datenverarbeitungsanlage als Instrument zur Entwicklung von Theorien, und zwar von Theorien zum Sprachverhalten.

Sprachbeschreibung und Sprachsimulation sind daher die zwei konstitutiven Gliederungsbereiche der CL, die unter sich vielfach verbunden sind. Diese Zweiteilung läßt sich auch in den einzelnen Anwendungsbereichen verfolgen.

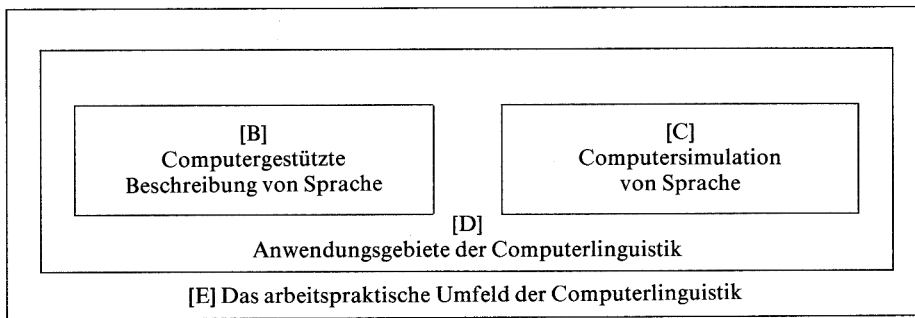
Die vorgelegte Systematik erhebt nicht den Anspruch auf Endgültigkeit; nach Meinung der Herausgeber reflektiert sie jedoch den gegenwärtigen Stand des Feldes zutreffend. Die Herausgeber hoffen, daß die in diesem Band vorgeschlagene Gliederung sich auch über den Rahmen des Handbuches hinaus als brauchbar und erweiterungsfähig erweisen wird.

### 5.2 Zur Gliederung des Handbuches im einzelnen

Die beiden zentralen Themenbereiche der Computerlinguistik, nämlich *Beschreibung* und *Simulation* von Sprache, werden in den Kapiteln IV—VIII des Handbuches systematisch dargestellt. Ihnen gehen einige epistemologische und historische Kapitel (I—III) voraus, in denen auch methodische Fragen behandelt werden, und ihnen folgen drei Kapitel, in denen die Anwendungsgebiete der Computerlinguistik vorgestellt werden (IX—XI). Das Handbuch schließt mit einem arbeitspraktischen Kapitel (XII) ab. Diese 12 Kapitel lassen sich demnach in fünf Hauptteile gliedern:

- [A] Entwicklung, Theorie und Methoden der Computerlinguistik (Kap. I—III)
- [B] Computergestützte Beschreibung von Sprache (Kap. IV—VI)
- [C] Computersimulation von Sprache (Kap. VII u. VIII)
- [D] Anwendungsgebiete der Computerlinguistik (IX—XI)
- [E] Das arbeitspraktische Umfeld der Computerlinguistik (XII).

In dem Hauptteil [A] wird die Einbindung der Computerlinguistik — wie bereits gesagt — in einen breiteren, historischen und epistemologischen, externen Rahmen versucht. Die interne Zuordnung der einzelnen Hauptteile kann folgenderweise veranschaulicht werden:



Der unterschiedliche Status des Wissens in den einzelnen Teilen wird besonders deutlich, wenn man die *Verwertungsdauer* des vermittelten Wissens betrachtet:

Während die linguistische Deskription [B] und Simulation [C] relativ stabiles Wissen darstellen, veraltet das Wissen in dem Anwendungsbereich [D] schneller. Noch dramatischer verändert sich das Wissen in dem arbeitspraktischen Umfeld [E].

Das Handbuch präsentiert eine Bibliographie von ca. 3700 Titeln zur Computerlinguistik, kompiliert aus den Literaturangaben der einzelnen Artikel. Sie wurde von Friedrich Wetter zusammengestellt, dem dafür an dieser Stelle herzlich gedankt sei. Obwohl die Herausgeber keine Vollständigkeit beanspruchen können, enthält das vorgelegte Handbuch die größte, umfassende, systematische Literatursammlung zur Computerlinguistik.

## 6. Computerlinguistik: Vision und Wirklichkeit

Es wäre sicherlich verfehlt zu behaupten, daß die Computerlinguistik in der Lage ist, eine komplette Beschreibung der Sprachverstehensprozesse zu liefern, die auch als Grundlage von Anwendungen der natürlichsprachlich orientierten Computeranwendungen dienen kann. Aber es ist klar geworden, daß die CL eine Grundlage liefern kann und teilweise bereits liefert und daß ohne eine Grundlagenforschung, die durch die CL abgedeckt ist, diese Ziele nicht erreichbar sind. Der gegenwärtige Einzug der CL in die akademische Lehre, der weltweit beobachtet werden kann, zeigt, daß die Tragweite der computerlinguistischen Grundlagenforschung auch außerhalb der engeren Fachgrenzen begriffen worden ist.

In diesem Zusammenhang ist die visionäre Analyse von Gerold Ungeheuer zu würdigen, der sich in einer Zeit, als der Zusammenhang zwischen Sprache und Computer noch durch Fehlvorstellungen und Vorurteile verdunkelt war, für die neue Disziplin eingesetzt hat. Er ist in zwei Hinsichten seiner Zeit voraus gewesen:

Er hat erstens in den Jahren — und diese Zeit ist noch gar nicht so weit von uns entfernt — als Geisteswissenschaftler den Computer noch ignorierten oder ablehnten, diesen Computer als Arbeitsinstrument der Sprachforschung erkannt und propagiert.

Er hat zweitens in den Jahren — und diese Zeit ist ebenfalls noch gar nicht so weit von uns entfernt — als Informatiker die Verarbeitung der natürlichsprachlich verschlüsselten Informationen noch als einen Präzedenzfall der allgemeinen Symbolverarbeitung betrachteten, erkannt, daß maschinelle Sprachverarbeitung nur in einem größeren, kommunikativen Rahmen bewältigt werden kann.

Das vorgelegte Handbuch zeigt, wie radikal sich die Welt seither geändert hat und wie zutreffend Ungeheuers Analyse von 1970 war.

Im März 1989

István S. Bátori (Koblenz)  
Winfried Lenders (Bonn)  
Wolfgang Putschke (Marburg)

## List of Maps/Verzeichnis der Karten

Karte 2.1: Forschungseinrichtungen der Computerlinguistik: Amerika	Compu- nach 18	Karte 44.4: Beispiel einer Lautkarte (Arbeitskarte aus dem KDSA 2, Karte 412: Bauern)	nach 560
Karte 2.2: Forschungseinrichtungen der Computerlinguistik: Australien	Compu- 20	Karte 44.5: Beispiel einer Wortkarte (aus WSAH, Karte 13)	564
Karte 2.3: Forschungseinrichtungen der Computerlinguistik: Europa mit Israel und Afrika	Compu- nach 20	Karte 44.6: Prüfpfadkarte; Prüfbezugspunkt 041065 (Ostfälisch)	nach 568
Karte 44.1: Beispiel einer Typenkarte (ALE QI, 72 Hafer)	nach 556	Karte 44.7: Relatives Kohärenzmittel auf der Grundlage des KDSA-Gesamtkorpus	nach 568
Karte 44.2: Beispiel einer Kombinationskarte (aus KDSA 1.1, Karte 188)	nach 556	Karte 44.8: Dialektometrische Kartierung der Belegort-Minima auf der Grundlage des KDSA-Abtaxenkonsonantismus	nach 568
Karte 44.3: Beispiel einer Similaritätstypenkarte	560		

## List of Figures/Verzeichnis der Abbildungen

Fig. 3.1: Example for a syntactic rule	61	maschinell erstellten Gebrauchswörterbüchern	233
Abb. 5.1: Paradigma Sprachdatenverarbeitung	82	Abb. 20.2: Hierarchie 'ontologischer' Merkmale, nach Schank 1969 (aus Lenders 1975)	235
Abb. 7.1: Betrachtete Baumstruktur	96	Abb. 20.3: Semantische Marker (nach Winograd 1972, 128)	235
Abb. 7.2: Speicherung bei einfacher Verkettung	97	Abb. 20.4: System semantischer Merkmale (nach Zelinsky-Wibbelt 1987, 188)	236
Abb. 7.3: Speicherung bei zweifacher Verkettung	97	Abb. 20.5: Semantisches Netzwerk (aus Lenders 1987)	238
Abb. 7.4: Analysematrix	99	Abb. 20.6: <i>Collocates</i> von <i>house</i> in absteigender Relevanzfolge (nach Berry-Rogghe 1973, 108)	240
Abb. 7.5: Folge der Aktionen	99	Abb. 20.7: Das semantische Feld 'Gebäude' (nach Berry-Rogghe 1973, 111)	240
Abb. 7.6: Erzeugte Baumstruktur	100	Fig. 21.1: Schematic state of the chart after morphological rewriting (Kay 1977, 141)	246
Abb. 10.1: Evaluierungskonzept FAS	124	Fig. 21.2: Letter tree with conflated shared initial parts for a small lexicon (Kay 1977, 162)	246
Abb. 14.1: Sonagramm	161	Fig. 21.3: Transducer for some English morpho-graphemic rules (Kay 1983 b, 100)	247
Abb. 14.2: Beispiele von Klassen sprachlicher Einheiten	165	Abb. 23.1: Verstehenstypologie	262
Abb. 15.1: Graphketten in Handschriften A. Dürers	170	Abb. 23.2: Ebenen des Verstehens (von Hahn 1978, 12)	266
Abb. 15.2: Die Repräsentation des mhd. < î > bei Krafft	172	Abb. 27.1: Assoziative Netzwerke (nach Quillian 1968) 'Die Schwalbe Sylvia besitzt ein Nest'	297
Abb. 15.3: Häufigkeitsverteilung der Wortformen mit den Graphketten < ie > und < ieh > 173—	174	Abb. 27.2: Die Vererbungshierarchie IS-A (nach Mylopoulos/Levesque 1983)	297
Abb. 17.1: Beispiel einer Analyse durch DERIVER 200		Abb. 27.3: Semantisches Netzwerk (Norman 1976, 10)	304
Abb. 17.2: Beispiel eines durch LEMMA 2 bearbeiteten Satzes (Satz 5 aus SFB 100 1980, 20)	203	Abb. 27.4: Architektur von symbolverarbeitenden Systemen	305
Fig. 18.1: Information structures in TG (standard model)	206	Abb. 27.5: Beispiel eines einfachen Kaskadierten Übergangsnetzwerkes mit Rekursion (nach Woods 1980)	307
Fig. 18.2: Information flow in process models	207	Abb. 27.6: Verarbeitung der Zeichenkette A B C (aus Christaller 1986, 117)	307
Fig. 18.3: Information structures in late TG (after Radford 1981)	208	Abb. 28.1: Compiler und Interpreter	313
Fig. 18.4: RTN for a simple sentence	208	Abb. 28.2: Die lineare Transfer-Architektur der MÜ	313
Fig. 18.5: ATN for a simple sentence	209	Abb. 28.3: Rekursionen	314
Fig. 18.6: ATN for passive sentence	209	Abb. 28.4: Die Blackboard-Architektur: Das Interlingua-Modell der MÜ	314
Fig. 18.7: DCG sentence rule	211	Abb. 28.5: Netzwerk-Architektur	314
Fig. 18.8: SLASH categories in a DCG	211	Abb. 28.6: Kaskadierung bei Hauenschild	315
Fig. 18.9: Passive metarule in GPSG	212	Abb. 28.7: Zweiphasen-Architektur für ein Parser-Generator-System	316
Fig. 18.10: Subject verb agreement in FUG	214	Abb. 29.1: An der Mensch-Maschine-Kommunikation beteiligte Wissensbereiche	321
Fig. 18.11: Active transitives in FUG	214	Abb. 29.2: Semantisches Netzwerk als Struktur von Knoten und Kanten	324
Fig. 18.12: Conflated rule in FUG	214	Abb. 29.3: Repräsentation des Satzes: <i>der Hund</i>	
Fig. 18.13: Typical LFG rules	215		
Fig. 18.14: LFG parse tree (c-structure)	216		
Fig. 18.15: The functional structure in the LFG	216		
Fig. 18.16: Information flow in LFG	216		
Fig. 19.1: The components of a signum	220		
Fig. 19.2: The basic factors in the explicative text interpretation	222		
Fig. 19.3: Basic notions of a semiotic text-theoretical framework	223		
Abb. 20.1: Natürlich-sprachliche Definitionen in			



<i>der den Mann gebissen hat, ist tollwütig</i> in einem semantischen Netz	326	Abb. 32.3: Prädikative Analyse mit Rücksetzen	383
Abb. 29.4: Die semantische Beschreibung des Satzes <i>Hans erschießt Maria</i> in der Notation von Schank	328	Abb. 32.4: Top-down Parser mit Vorausschautabelle	385
Abb. 29.5: Repräsentation eines Satzes in MEMOD (nach Norman/Rumelhart 1978, 66)	329	Abb. 32.5: Shift-reduce Parser mit Rücksetzen	387
Abb. 29.6: Beschreibung eines Besuches in der Sprache KRL (nach Bobrow/Winograd 1977)	332	Abb. 32.6: Bottom-up Parser Breite zuerst	389
Fig. 30.1: One to one relation between a word and its analyses	339	Abb. 32.7: Tabellengesteuerter Shift-reduce Parser	391
Fig. 30.2: One to many relation between a word and its category analyses	339	Abb. 32.8 a: Nicht-deterministische Grammatik und Steuertabelle	393
Fig. 30.3: One to many relation between a word and its subcategory analyses	339	Abb. 32.8 b: Tabellengesteuerter Shift-reduce Parser für nicht-deterministische Grammatiken	394
Fig. 30.4: Many to one relation between word and syntactic analysis	340	Abb. 32.9: Parsing mit dem Algorithmus von Earley	397
Fig. 30.5: One to many relation between a word and its semantic analyses	340	Abb. 32.10: Bottom-up Parser nach Cocke, Younger und Kasami	401
Fig. 30.6: Separate category and subcategory information	343	Abb. 32.11: Verallgemeinerter Bottom-up Parser mit Teilergebnistabelle	404
Fig. 30.7: Category and subcategory information as features	343	Abb. 32.12: Erkennen mit Finitem Übergangsnetzwerk (FTN)	408
Fig. 30.8: Complement slots in lexical entries	343	Abb. 32.13: Parsing mit rekursiven Übergangsnetzwerken (RTN)	411
Fig. 30.9: Entries from a full entry lexicon	343	Abb. 32.14 a: Erweiterte Übergangsnetzwerke (ATN)	415
Fig. 30.10: Regular and irregular forms in DIAGRAMM	344	Abb. 32.14 b: Ein ATN-Programm	416—418
Fig. 30.11: Regular and irregular forms in PATR-II	345	Abb. 32.14 c: Das Lexikon zum ATN-Programm	418—419
Fig. 30.12: Regular and irregular forms in the BBN Context Free Grammar	345	Abb. 32.14 d: ATN-Abarbeitungsbeispiel	419—420
Fig. 30.13: Sharing information among lexical entries for irregular forms	345	Abb. 32.15: Parser für Kategorialgrammatiken	423
Fig. 30.14: Sharing information among entries in a full entry lexicon	346	Abb. 32.16 a: Lexikalisierte Dependenzgrammatik	428
Fig. 30.15: Collocations entered directly in the lexicon	346	Abb. 32.16 b: Teilergebnistabelle des Parsers für Dependenzgrammatiken	429—430
Fig. 30.16: Collocations contained in another lexical entry	347	Fig. 34.1: Case relations in English (from Nagao/Tsujii/Nakamura 1985)	442—443
Fig. 30.17: Collocations dependent on other lexical entries	347	Fig. 34.2: Winograd's blocks world	444
Abb. 31.1: Interpretierender Parser	350	Fig. 34.3: Dictionary definitions for 'cube' and 'contain' in Winograd (1973)	445
Abb. 31.2: Prozeduraler Parser	350	Fig. 34.4: The form of syntactic representation in Winograd (1973)	445
Abb. 31.3: Compilierter Parser	351	Fig. 34.5: Word senses in Wilks (1978 b)	446
Abb. 31.4: Baumgraph	365	Fig. 34.6: The template for <i>John shut the door</i> in Wilks (1983 a)	447
Abb. 31.5: Liste	365	Fig. 34.7: The form of CD representation in Schank/Rieger (1974)	448
Abb. 31.6: Klammersausdrücke	365	Fig. 34.8: Parsing of the sentence <i>John gave Mary a beating</i> in Riesbeck (1975)	449
Abb. 31.7: Gesuchter Strukturbaum	371	Fig. 34.9: The form of ATN representation in Woods/Bates/Brown et al. (1976)	451
Abb. 31.8: Top-down Analyse	371	Fig. 34.10: The representation of feature hierarchies by network in Collins/Quillian (1969)	452
Abb. 31.9: Bottom-up Analyse	371	Fig. 34.11: The semantic representation in Quillian (1967)	452
Abb. 31.10: Tiefe zuerst	372	Fig. 34.12: Different types of links in networks	453
Abb. 31.11: Breite zuerst	372	Fig. 34.13: The representation of 'smoking' in the FRAME used by Wilks (1983)	456
Abb. 31.12: Chart	373		
Abb. 32.1: Top-down Parser mit Rücksetzen	379		
Abb. 32.2: Top-down Parser mit paralleler Abarbeitung	382		

Fig. 34.14: Issue skeletons used for the description of terrorism by de Jong (1979 a)	456	Fig. 42.2: Table of dissimilarities	536
Fig. 34.15: Sketchy script for 'nationalize'	456	Fig. 42.3: Four points in a tree	540
Fig. 36.1: Language generation systems	472	Fig. 42.4: Configurations of four points	541
Fig. 36.2: Central issues addressed in various language generation projects	473	Fig. 43.1: Words evoking the theme of <i>l'incertitude</i>	547
Fig. 36.3: Knowledge and information sources needed in language generators	475	Fig. 43.2: Co-occurrences with <i>travailleurs</i>	548
Abb. 37.1: Das Dreiphasenmodell der Wissenschaft (nach Andersson 1978)	483	Fig. 43.3: Frequencies from <i>Le Rouge et le noir</i>	550
Abb. 37.2: Kritizistisches Modell der Wissenschaft (nach Andersson 1978, 66)	484	Fig. 43.4: Standard deviates and degree of probability	551
Fig. 38.1: Word Index to the Gothic Bible (de Tollenaere/Jones 1976)	490	Abb. 44.1: Konzeptioneller Entwurf der automatischen Sprachkartographie	554
Fig. 38.2: Concordance to Beowulf (hand) (Cook 1911)	491	Abb. 44.2: Logische Form der Datenmatrix	556
Fig. 38.3: Concordance to Beowulf (computer) (Bessinger/Smith 1969)	491	Abb. 44.3: Übersicht über die Lautabstände des Distanzverfahrens (aus Naumann 1977, 201—202)	558
Fig. 38.4: Concordance to the Bible Cruden, 17	491	Abb. 44.4: Beispiel einer computativen Klassifikation mit Hilfe des Distanzverfahrens (aus Naumann 1977, 166)	558
Fig. 38.5: Concordance to Shakespeare (Bartlett 1889)	492	Abb. 44.5: Beispiel einer computativen Klassifikation mit Hilfe des Substitutions- bzw. Reduktionsverfahrens (ALE QI, 32 Kupfer, germanisches Departement)	559
Fig. 38.6: Concordance to Matthew Arnold (Parrish 1959)	492	Abb. 44.6: Vorschlag für eine systematische Symbolkonstruktion im Rahmen des Europäischen Sprachatlas	561
Fig. 38.7: Concordance to Macbeth (Howard-Hill 1971)	493	Abb. 44.7: Beispiel einer systematischen Symbolkonstruktion (ALE QI, 496 Papier, Ausschnitt)	562
Fig. 38.8: Concordance to Shakespeare (Spevack 1969)	494	Abb. 44.8: Beispiele einer sprachkartographischen Flächendarstellung (aus Händler/Naumann 1976, Abb. 10—13)	563
Fig. 38.9: Intermediate output for lemmatizing	495	Abb. 44.9: Überblick über die konzeptionelle Weiterentwicklung der automatischen Sprachkartographie	565
Fig. 38.10: Concordance to Faust I (Chisholm/Sondrup 1986)	496	Abb. 44.10: Dialektrometrische Matrizen (nach Goebel 1982, 24)	567
Fig. 38.11: Concordance to Cummings' Poetry (McBride/Parrish 1987)	496	Abb. 44.11: Nhd. Anteil in den deutschen Dialekten auf der Grundlage des KDSA-Gesamtkonsonantismus in Stereogrammdarstellung (Werte ohne Klasseneinteilung) aus südwestlicher Perspektive	569
Fig. 38.12: Concordance to Livy (Packard 1968)	498	Abb. 44.12: Nhd. Anteil in den deutschen Dialekten auf der Grundlage des Gesamtkorpus in Stereogramm-Darstellung (in 6facher Klasseneinteilung) aus südwestlicher Perspektive	569
Fig. 38.13: Concordance to Freud (Guttman/Jones/Parrish 1980)	499	Fig. 47.1: Processes involved in speech recognition	588
Fig. 38.14: Frequency List to the Gothic Bible (de Tollenaere/Jones 1976)	500	Fig. 47.2: Illustration of a finite state machine (Markov model)	591
Fig. 38.15: Reverse Word List to the Gothic Bible (de Tollenaere/Jones 1976)	501	Fig. 47.3: Typical structure of a system for isolated word recognition	592
Fig. 38.16: WordCruncher View Reference List, forms of <i>heilig</i> in Goethe's Faust	507	Fig. 47.4: Illustration of connected word recognition	594
Fig. 38.17: WordCruncher View Reference List, combination of <i>Gott</i> etc. and <i>Mensch</i> etc. in Goethe's Faust	508	Fig. 47.5: Search procedure continuous speech recognition	596
Abb. 40.1: Aufbau der Belegdaten	522	Fig. 47.6: Example of a finite state network as language model (after Jelinek 1976)	597
Abb. 40.2: Ausdruck zur Auswertung der Substantivdaten	523	Abb. 48.1: Sprachsynthese (Ablaufdiagramm)	601
Abb. 40.3: Ausdruck einer Übersicht über die automatisch ermittelte Belegung in den einzelnen grammatischen Positionen	525	Abb. 48.2: Detaildiagramm „Methoden der Sprachsynthese“	602
Abb. 40.4: Entscheidungsbaum der automatischen Segmentierung der Substantive (aus Berg/Frauenknecht 1982, 128)	527	Abb. 48.3: Beispiel für Sprachsynthese nach dem	
Fig. 42.1: Table of reading numbers	536		

Verfahren Concept-to-speech (nach Young/Fallside 1969, 688)	605	Fig. 52.4: A naive view of relationships between words and concepts	659
Abb. 48.4: Komponenten von Sprachsynthesensystemen	606	Fig. 52.5: Terminological words and concepts	659
Abb. 48.5: Zur Umsetzung schriftsprachlicher und optischer Eigenschaften von Texten in lautsprachliche/akustische Formen	607	Fig. 52.6: Vagueness of word meanings	661
Abb. 48.6: Geräte zur Sprachsynthese (Stand Sommer 1985, ergänzt 1986)	608—	Fig. 52.7: Understanding and surface determination in future MT Systems	663
Abb. 48.7: Kosten heutiger Sprachsyntheseanwendungen (nach Sclater 1983)	610	Abb. 57.1: Grundstruktur eines Frage-Antwort-Systems	709
Abb. 48.8: Zusammenhang zwischen Wortbetonung und Satzbetonung bei der Sprachsynthese	613	Abb. 57.2: Grundstruktur eines Interface-Systems	709
Abb. 48.9: Sprachausgabeanwendungen	617	Abb. 57.3: Die zentrale Stellung von Wissensrepräsentationen	710
Abb. 48.10: Synthetisierte gesprochene Sprachen	619	Abb. 57.4: Der Frage-Beantwortungsprozeß	712
Fig. 49.1: The most common model for machine translation systems	623	Fig. 58.1: The basic components of question answering systems	722
Fig. 50.1: Overview of the operational Japanese MT systems	641	Fig. 58.2: USL — Semantic tree	725
Fig. 51.1: The architecture of English-Russian MT in AMPAR (Marčuk 1983, 93)	646	Fig. 58.3: Parsing with transformational grammar after Winograd (1983)	726
Fig. 51.2: The architecture of German-Russian MT in NERPA (Marčuk 1983, 93)	647	Fig. 58.4: Scene from the block world	733
Fig. 51.3: Scheme of general design of MT with translational equivalents (Marčuk 1982, 106)	648	Fig. 58.5: Flow charts for noun group — according to Rubin (1973)	735
Fig. 51.4: The architecture of French-to-Russian MT in FRAP (Leont'eva/Nikogosov 1982, 144—145)	649	Fig. 58.6: A summary of the question answering systems reviewed	738
Fig. 51.5: Machine translation systems in East Europe	651	Fig. 60.1: Linear programs (Zettersten 1986, 22—23)	745
Fig. 52.1: A naive schematic view of translation which is often used but quite misleading	654	Fig. 60.2: Author-initiated branching	746
Fig. 52.2: A schematic view of future MT systems	656	Fig. 60.3: Student-initiated branching	746
Fig. 52.3: A simplified framework of an NLU	657	Fig. 60.4: The COMMENTATOR world	747
		Fig. 60.5: The text production model of COMMENTATOR	748
		Fig. 60.6: BRAINLEARN: A new learning system for English, German, French, etc. by Arne Zettersten and Mats Jacobson	750
		Abb. 62.1: ISO 6937, 8-Bitcode, Lateinische Alphabet- und Nichtalphabet-Zeichen	763