die Abbildungs-Problematik in grundsätzlich anderer Weise aktuell. Üblicherweise kann bei prädikativer Programmierung nicht zwischen Daten und Anweisungen unterschieden werden. Beide Objektklassen bedürfen der Formulierung von Aussagen über sie und werden durch diese definiert. Beispielsweise ist es mit gewissen Varianten von PROLOG möglich, Datenbanken und die zugehörigen Manipulationsverfahren mit einem gemeinsamen Anweisungsformat zu spezifizieren.

Deshalb ist zum Abschluß des Beitrages einschränkend zu konstatieren, daß die vorliegenden Erörterungen zunächst für eine linguistische Datenverarbeitung auf der Grundlage prozeduraler Programmierung Gültigkeit haben. Zu den prozeduralen Sprachen zählen etwa FORTRAN, PASCAL, COMSKEE und ADA sowie die überwiegende Mehrzahl der sonstigen gebräuchlichen Sprachen. Für nichtprozedurale Sprachen und Systeme gelten andere Kriterien, die vermutlich in den nächsten Jahren Gegenstand der weiteren Forschung sein werden. Mit dem Hinweis auf die Aktualität dieser Fragen soll allerdings nicht angedeutet werden, daß sich das prädikative Paradigma in einem bestimmten Anwendungsbereich auf Dauer als einziges durchsetzen könne. Es dürfte eher zu einer fruchtbaren Debatte um Vor- und Nachteile der unterschiedlichen Ansätze kommen, als deren Resultat bestimmten Teilen der prozeduralen Programmierung anhaltende Bedeutung zuerkannt werden muß.

6. Literatur (in Auswahl)

S. Baase 1978 · E. Bertsch 1985 · E. Bertsch/A. Mueller-von Brochowski 1979 · G. Booch 1983 · N. Chomsky 1959 · D. R. Dowty/L. Karttunen/A. M. Zwicky 1985 · J. Earley 1970 a · Z. Galil/J. Seiferas 1983 · S. L. Graham/M. A. Harrison/W. L. Ruzzo 1980 · W. Hansen 1969 · M. J. R. Keen/G. Williams 1985 · D. E. Knuth 1971 · H. Maurer 1969 · E. Y. Shapiro 1983 · E. Y. Shapiro/A. Takeuchi 1983 · H. J. Schneider 1975 · F. Stetter 1981 · H. Wedekind 1970 · P. H. Winston/B. K. P. Horn 1981 · N. Wirth 1982 · D. H. Younger 1967.

Eberhard Bertsch, Bochum
(Bundesrepublik Deutschland)

8. Application and Research — Mutual Side Effects

1. Research, Applications and Honesty
2. Apply What?
2.1. Uniqueness of Natural Language
2.2. Artificiality in Natural Languages
2.3. Incompleteness of Natural Language Descriptions
2.4. User-Tolerance of Natural Languages
2.5. Vagueness in Natural Languages
2.6. Natural Language Are Languages in the Making
2.7. Perfect Translations Do Not Exist
2.8. More Than Transmitting Information
3. Apply How?
4. Apply to What?
4.1. Designing Procedures for Reading, Writing and Interaction
4.2. Designing Tools for Language Learning and Teaching
4.3. Designing Languages
4.4. Observing Language Use and Language Users
5. Literature (selected)

1. Research, Application and Honesty

Many Research and Development projects in natural language processing are improductive: they are not ventursome enough to yield new knowledge, and the knowledge they build on is not established safely enough to give robust tools which can be used for practical purposes. When attacked because a project produces only trivial knowledge, the planners refer to production needs: After all, they say, we have users to serve and cannot play around irresponsibly. And when users complain that the system does not perform properly, they are dismissed with the remark that building up knowhow is more important than immediate production. Probably this risk for risk evasion is inherent in the very concept of Research and Development. It would be much more appropri-

ate to subsume *Production and Development* under one heading, and to speak of P-and-D instead of R-and-D. Development can be seen as an activitity, albeit one with delayed effect, towards production. Research cannot.

Both sides would profit from recognizing that research and application are different not in degree but in kind. They have different goals. Sometimes theoretically meagre research is classed as 'applied', but a study does not become practical only because it lacks theoretical depth — nor do unpractical attempts, often with an unpolite phrase said to be of "merely academic interest", necessarily possess theoretical value. On the contrary: nothing is so practical as a good theory, and there are few better theoretical challenges than a precise practical problem, empirically unsolvable by mere practicalities.

Research does not aim at production, now or ever. It aims at knowledge and understanding. CL as such is not linguistics applied to the use of computers or computer science applied to linguistics (a misconception rejected by, e. g., Bátori 1977 a). It is not tantamount to linguistic data processing to process language data, for theoretical or practical purposes. The core concept of CL is not 'computer' but 'computation', the old idea of a 'calculus', a rule-bound play on symbols as with pebbles. CL explores what is computable in language. Ultimately it seeks an answer to the old question: What is mechanical in us?

This question is ancient, but the advent of computers has made it inescapable and given new angles to look at it. It would have been worth while to construct computers if only for the purpose of better formulating the questions of CL. How absurd the suggestion that applications to, say, office automation should be a justification for us to study the most human of all behaviour, the use of language and the shaping and reshaping of language!

One way of studying language, or ourselves as creatures who interact and think using language, is to model our linguistic capabilities in computing machinery. Whereas engineering is concerned with machines acting instead of men, CL studies man as a machine. Whenever we can reduce human behaviour to mechanical procedures, we gather urgent, not seldom painful knowledge about ourselves; when we fail, and recognize why, we may learn even more. It is not only in thermodynamics that the great failures mark the fundamental advances.

To take only one example: we have learnt more during the last thirty years about translation than during centuries before because we now know that manipulation of syntax and lexicon alone will produce intelligible or otherwise useable output only under special circumstances; we are still struggling to determine exactly which these circumstances are. This is true although grammatical analysis has indeed reached a precision one could hardly dream of in the optimistic fifties, computer storage and speed have eliminated all fear that the machines could not accommodate any lexicon linguists could ever manually compile, and machines now do produce impressive translations in some situations (Cf. art. 50—52). We have perceived fundamental obstacles in a clearer form than any sceptic could before. We have gained new respect, if you like, for the human performance even when the translation task seems trivial and mechanical. The very success of automatic language processing has established more fundamental failures. And the new insights gained have immensely higher value than hundred-per-cent adequate black-box translating robots could have had. After all, those could at best have given us more and cheaper translations (Karlgren 1981).

Being mutual side effects, research and application can profitably exploit each other, if and only if they honestly state their goals. To pick an example from my own experience, we linguists offer solutions to the production of adequately abbreviated Braille writing on paper or tactile 'screens' for the blind, who thus profit from linguistic research to get less antiquated reading materials (House 1970 b). At the same time, people who are accustomed to using fingers instead of eyes for reading help us to understand text perception better. As scholars we are glad that there exist informants who can report on other linguistic channels than the ordinary ones. The blind can exploit linguistics and the linguists can exploit the blind. Funding agencies, planners, and even scholars are often so brainwashed by their own apologetic exercises in non-scientific terms that they find such blunt statements cynical and frivolous. In contradistinction, blind people and their organizations have little difficulty in perceiving and appreciating their two roles as receivers and givers in relation to research. Scholars are urged by sponsors, adminis-

trations, universities and presumed public opinion to justify their activity by reference to alleged future applications. When they acquiesce, they trivialize their own real pursuit and perform poorly as consultants. A few times too often, large scientific projects have remained futile and failed to deliver results. Thus we have seen projects in many countries pretending to produce translation systems but being only excuse for theoretical study of language. In later times, documentation and expert systems have been the targets of similar half-hearted attempts.

As a result, we still have few linguistics-based useful tools. Linguistic spin-offs have often been poor, and the respect for scholars has been eroded. Hence, ironically, the prospects have diminished for research support. Thus, translation aids which could have been operative decades ago have been left undesigned, because they were not fascinating enough for scholars but yet beyond what could be delivered by mere computer specialists and office automation industry (Kay 1980 b). And for the same reasons, the commonly used word-processors of today work essentially as the typewriters which met the first generation of office girls, with some much-boasted mechanisms to support the typists in doing corrections, amendments, page layout up to and including footnotes, and maintaining small archives, whereas next-to-nothing is offered to support authoring; currently used information retrieval systems are still primitive, featuring little beyond boolean operations on word occurrences, without linguistic refinement: progress in retrieval is due to more powerful computer facilities, permitting mass text processing even to individual and private users, and the 'full-text' approach is as yet most often a brute-force way of replacing intelligent analysis by using more data, on a lower level of sophistication than in 1950 and 1960 when considerable amounts of semantics were squeezed into classification codes, descriptors and other text amplifications (Karlgren/ Walker 1980).

2. Apply What?

If someone programs a computer to treat an English text to format it or interpret it as a command, that project is not necessarily a case of applied linguistics. To be, it is neither enough nor necessary that the objects processed are natural language data. However, the term 'applied' does presuppose that something is being applied. But what is being applied in computational linguistics?

Non-linguists have a tendency to see linguistics as a much narrower field than linguists do. They ask linguists about too little, too late and about the wrong issues. This is unpractical.

Let us loosely define linguistics for our present purpose as the knowledge peculiar to linguists and a linguist as someone who has spent a considerable part of his life in scientific study of natural languages. CL, accordingly, is the knowledge derived from scientific study of computational aspects of language. What wisdom, then, have linguists acquired? What insights, beyond encyclopedic details on existing languages, is there to apply?

The following observations about natural languages, we have found, are liable to be overlooked by persons with an engineering or artificial intelligence background.

2.1. Uniqueness of Natural Language

Natural Language, in singular, does not exist. Or if it does, it is a highly artificial product, which features English spelling and which is boasted to be acceptable and understandable to laymen without special training.

It is true that linguists have laid bare a common basic design in *all* human languages. The recognition of such universal design is something linguists have laboriously arrived at, a major achievement of later decades (for overviews see, e. g., Greenberg 1966; Lenneberg 1966; Wunderlich 1974, p. 93 ff.). But in many important ways languages do differ, from place to place, person to person and profession to profession, and the differences, though more superficial than one might be inclined to believe, are not negligible.

If you are told that you can use just plain English to interrogate a data base, how can you know what kind of English the system is going to accept? How can you know — how could you be told? — what meaning the thing is going to assign to your statements? By referring to Natural Language the designer evades the problem of explaining to the user what he may do and how his commands will be understood. What the user is in fact offered is (an unspecified superset *of*) a rather arbitrary unspecified subset of the special-purpose English he would have used if there

had been an English-speaking person at the other end of the line.

2.2. Artificiality in Natural Languages

Natural languages are not natural. The idea of Nature as opposed to Art dates back to antiquity. A recent variant is the concept of artificial languages, such as programming languages or query languages, contrasting with natural languages, or Natural Language, which have grown (allegedly) untouched by human hand. This view is at variance with obvious facts: who could assert that the English of medical patient records, law reports, or Letters of Patent has developed spontaneously and naturally? All the languages for special purposes — 'Fachsprachen' — which computer-supported information processing is typically concerned with are highly, though not exhaustively, controlled by very deliberate regulation of terminology, phraseology, style and composition. Conversely, so-called artificial languages as actually used by humans, display regularities which have developed during usage without conscious decisions and which are not laid down in formal definitions. A parser of human-produced input in an artificial language, therefore, can do better than run to rules. Applying linguistic (sic!) analysis it can, with a certain risk of failure, recognize intended messages; an early example of this technique is the DWIM *(Do What I Mean and not what I say)* package for Interlisp. In such cases, the parser assumes that the explicit rules are disobeyed and that the user is obedient to general regularities of which he as well as the language inventor are happily unaware.) Instead of debating for or against natural language input and output, we should ask ourselves what features in existing languages we want to take over into languages to be invented, i. e., what kind of naturalness is desired for a particular application.

2.3. Incompleteness of Natural Language Descriptions

Natural languages are incompletely mapped. An artificial language is in some sense always completely known. The map is made before the voyage and is, by definition, right. Natural languages are complex, and we keep on disclosing more and more complexities. There is no reason to hope that our description will ever be exhaustive. Linguists do not share the expectation that could we only complement the lexicon, expand the text base, refine the rules and perfect the the semantic representation — i. e., finalize the thesaurus — we would have practically everything under control. (The hypothesis, that natural language provides only labels for language-independent pre-existing concepts has long since been abandoned by linguists. The opposite is argued by, e. g., Humboldt, W. 1836; Cassirer 1923—31; and in extreme and popularized form Whorf 1956). We must keep alive the money-forgers doubt: there might be more to it.

The proper conclusion of this awareness of indeterminacy is not seldom to simplify the system so that it can live with incomplete knowledge, rather than take into account a few more observations. For instance, rather than extend a vocabulary, we may spend our efforts on making the system react adequately when new words or constructs are encountered.

2.4. User-Tolerance of Natural Languages

Natural languages are tolerant towards sloppy users: that is what they were designed for — except that, of course, their basic physiology was never deliberately designed at all.

Natural languages strike a phantastic balance between economy and safety: we can, as everyone knows, cut out sentences, phrases or words, replace cumbersome words by pronouns or other pro-words, reduce pronunciation or handwriting to a mere hint, etc., and yet be understood by listeners or readers who lend us a distraught ear or eye in a noisy environment.

A modern theory to account for the old observations that humans so amazingly seldom fail to identify heavily reduced and badly distorted utterances and texts came with the advent of Shannon's mathematical theory for the transmission (sic!) of information, later called merely information theory, a term often misinterpreted: Shannon 1949; Shannon/Weaver 1948; Cherry 1957; Meyer-Eppler 1969. (For early very systematic observations of what was later called 'redundancy' on all levels, including syntactic and 'textual' levels, so-called 'Prädikatkürzungen', German-readers should see Gabelsberger 1843, commented in Karlgren 1973).

It is one of the wonders of language that it caters for our laziness — not even asking us to learn explicit abbreviation rules. The options in some programming languages to

allow concise and verbose forms are very modest attempts to exploit redundancy.

2.5. Vagueness in Natural Languages

Natural language expressions are often vague. In addition to ambiguous and imprecise statements — with more than one interpretation or with an intended range of interpretations, respectively — there are expressions for which neither speaker or listerner could produce an exact interpretation (Pinkal 1983). Logicians and engineers are prone to believe that this is always a matter of incomplete analysis or of incongruence between the communicants' semantic systems. But a frequent if not a normal case is that the expressions are inherently vague. It is not only that we have not conveyed our code yet to our next man. There is no exact meaning of *new car, nice girl, great expectations.* This vagueness — which is a normal state of affairs in everyday thinking — is the semantic counterpart of sloppy writing and speaking. It is permitted and does *not* create chaos!

It is in recent years, having formal tools for the precise description of complex meanings — including modality, belief structures and the like, which were traditionally outside the scope of all formalisms — that the intrinsic vagueness in natural language expressions becomes obvious: we have come to a point where we do not *want* our description to be more precise. Earlier linguists have tended to attribute the range of interpretations to not-yet-known variations between older and younger usage or between geographical areas; and in recent years, as long as the logical explications remained inadequate, semanticists would typically direct their main efforts towards achieving higher resolution. Good lexicologists, however, have long been conscious of the requirement to be suitably precise and to curb their desire to make logically exact definitions. — A fair statement of the situation is given by Lieb (1983, 212): "The topic of vagueness is very much at the centre of current semantic interest. I must confess that I am not satisfied with any approach to vagueness that has been suggested; nor can I offer a satisfactory alternative. I do believe, though, that my conception of lexical meaning does not preclude a satisfactory treatment." See also Lieb 1980. Cf. further Karlgren 1979, 182—187, 188 ff.; Pinkal 1983; Lakoff 1972; and, with regard to the consequence for information retrieval, Karlgren 1977.

It should be noted that the vagueness referred to has little in common with the fuzziness defined by Zadeh 1975, and in his later work; cf. (1982, 425 ff.). Zadeh's interesting model, where class membership is a stochastical variable, accounts for indeterminacy, but characteristic for human communication is not only that meaning may be indeterminate at any given point of time but also that vagueness can be successively reduced (or increased) through dialogue interaction and text progression. Zadeh's model is as static as the deterministic models in the sense that the language system remains unchanged.

By way of dialogue, higher clarity can be achieved in most cases, and normally humans do arrive at the level of (cognitive) understanding they require. This is another wonder of language. A challenge to engineers if you please, like the bird's flight was to Leonardo da Vinci.

The dialogue-based comprehension procedure in human discourse can not be replaced by an analysis based on ever-so-large static lexica or pre-established semantical representation (see 28). One cannot define oneself out of the necessity to cope with undefined expressions. — Another way to put it: the linguistic operations during a dialogue have as one operand the transmitted message and as one operand the language the participants are sharing. The speakers are creating, or modifying, for their immediate purpose the very tool they are using (Enqvist 1987). Essentially the same mechanisms are at work in language learning, language change and the comprehension of a coherent text. And this is true not only in literature but in any text, even in the most disciplined and 'formal' legal or mathematical prose. (Kiefer 1983; Allen/Perrault 1980; Wunderlich 1974, 225 ff.; Ungeheuer 1977, 1974)

For better and for worse, this is a major kind of naturalness which is absent in constructed languages. Semantical rigidity has a value of its own in some cases — that is why such artificial languages as predicate calculus were designed in the first place — but in other situations it makes them hopelessly unhospitable. Whatever efforts are spent on surface phenomena like spelling, word order, or inflection, a language unable to live with inherent vagueness will remain a set of formulae in disguise, incapable of true dialogue. Dialogue is more than exchange of utterances

in a semantically stable language. Exactly what more? We do not know yet — applications have to wait for research.

2.6. Natural Languages Are Languages in the Making

This observation is closely related to the previous statement. It is always possible to extend a natural language by (re)defining terms or merely by suggestive usage. Vagueness is one price paid for the incredible flexibility which makes it possible to go from nursery to high-tech laboratories or big business with almost unnoticeable little adjustments. Or from one century's pre-industrial life to information society. Humans have the habit of language shift: they accept it, understand it, and do it, without signalling a warning beforehand. Every language which is 'closed', even thought it can be extended by explicit definitions in a meta-language, will be felt as stifling or, if users do not realize its restraints, will lead to misunderstandings. (A survey, with rich bibliography, is given in Schröder P./Steger 1981. A survey and a synthesis of research in the field is to be found in Enqvist, 1987. For traditional views cf. Paul 1909; Wyld 1932. In well-needed efforts to sharpen linguistic methodology, the proper appreciation of vagueness was lost, or lost sight of. The sharp logical distinctions between describing states and describing transitions made it harder to view the interrelations; cf. Koerner 1973. — For more recent introductory literature, see Anttila 1972, and Donato 1980.)

2.7. Perfect Translations Do Not Exist

Serveral definitions of an appropriate translation have been offered; they share the property that, if taken seriously, no earthly translation can make a full score. The intrinsic intranslatability of natural languages was noted already by Leibniz and by Locke 1690; cf. Fraser 1894; Aarsleff 1975, p. 397.

This is true for translations from one natural language into another and it is true for translations between a natural language and algebraic or network representations. No matter how clever the mapping procedure, how large our fact bases, and how exact our dictionaries, there is no such thing as the correct explication of all and only what was stated in one language in another's representation.

Interfaces between humans using natural languages and machines using theirs will therefore necessarily introduce errors — tolerable, or at best removable during dialogue. One cannot isolate the language problems into an interface module and then forget about them.

2.8. More Than Transmitting Information

Natural languages have other functions besides the function to transmit, store, and structure information (Bühler 1934). Even the most factual statements express other functions which are as important to human beings as the transfer of information. Establishing groups of 'us' and 'them' to take an old example. The classical telegraphy-inspired kind of transmitter-channel-receiver model for language disregards — as models should — every other aspect except one, that of cost and safety of transmission. Linguists are beginning to rise above the Shannon-Weaver model of communication and to state a few general things about dialogue and polylogue behaviour.

These insights are highly relevant for systems design, even when the systems have factual information transfer as their sole purpose. Different people will adopt and retain different terms for the 'same' concept because they see the world differently, and want to see it differently (Codd 1974). This imposes a limit to standardization: The engineering, accounting and sales department of the same company will use different words for things and events which in objective terms are the same. If forced to use a standardized vocabulary they will find it unnatural and disgusting — and sooner or later give different interpretations to messages they send or receive. A good deal of well-meaning term bank efforts build and fail on the assumption that lacking conformity is due to the residue of terms not yet defined or to the untrained users' ignorance of standard terminology.

These major insights should be kept in mind in language processing. — A further point deserves mention, the most salient one: these devices work, in spite of the 'idiosyncracies' we find in them all. Or because of them?

3. Apply How?

How should linguistic knowledge be brought into action in applications? Certainly not via scholars producing commercial software.

There are three roles for the linguist:
— an *instructive* role: act as a broker of tools and results, which are there for to be taken. Existing language processing methodware can be put to work; resources spent on projects to reinvent them can be saved for other purposes (Karlgren 1986; Walker/Karlgren/ Kay 1977).
— a *constructive* role: to suggest tools and other facilities which could be produced, if research resources were directed towards the crucial areas.
— a *destructive* role: to inhibit or discontinue development projects for which the chances of success are nil. In a number of areas, allegedly practical, people spend time and energy on attempts which were known from the very beginning due to experience and theoretical considerations, to be abortive. Thus, much of the efforts given to produce limited unambiguous and semantically stable vocabularies or thesauri for information retrieval are necessary failures, not as non-linguists tend to believe because the resources have been too small to make them complete or to be specific enough in defining hierarchial or other relations but for more fundamental reasons. The idea of the ideal conceptual language, as linguists can tell their documentalist friends, has been tried before and given up — a few centuries ago. There is no such thing as a perfectly exact and timelessly stable language to be discovered or constructed for reasons known (as discussed in 2.4, 2.5 and 2.6). Cf. the well known attempt to construct an ideal language by Leibniz (Dutens (ed.) 1768; Loemker (ed.) 1969; Lenders 1976; Arndt 1967; Verburg 1951, particularly p. 220 ff. with comparisons with Locke). A less known example is Wilkins 1668, commented upon in Verburg 1951, 297 ff. and in Karlgren 1978 (German translation Karlgren 1979).

4. Apply to What?

Four situations can be identified where CL can be applied: 1. Using language, 2. learning to use language, 3. constructing (or reconstructing) language and 4. observing language users as they produce or comprehend language. Accordingly the applications can be grouped together under the following headings:
— Designing procedures for reading, writing and interacting
— Designing tools for language learning and teaching
— Designing languages
— Observing language use and language users.

The terminology reflects the user's perspective and it calls putting information into the machine 'writing' and displaying information by the machine 'reading', leaving the inverse terminology (the system's view) to other occasions when we discuss computer-internal processes. We shall also subsume speaking and listening under writing and reading, respectively, since we deem the distinction between written and oral communication to be of minor importance for the over-all view.

4.1. Designing Procedures for Reading, Writing, and Interaction

4.1.1. Writing

Writing in the most literal sense is an activity which has been almost completely neglected by scholars and commerce alike. *Keyboards* have remained essentially unchanged during the electronic revolution, and *typing* shortcuts and shorthand have been almost entirely overlooked. Substantial results of importance to nearly everyone could be achieved relatively easily if experts cared to condescend to such trivialities and investors showed the commercial audacity to challenge some basic conventions. The much harder problem of *speech recognition* has received more scholarly attention and almost unlimited funding. Although practical results have emerged in later years for specific situations, the attempts at replacing typing by speech more generally have largely failed during the era of computer utilization we have witnessed. This is not surprising since speech recognition involves perception and comprehension on many levels which are incompletely known and poorly understood. Typically, successful systems are contained in ambitious *speech understanding* schemes, e. g., for recognition of commands, whereas the seemingly simpler 'phoneme detector' is no longer even a dream. Taking dictation in an office, every typist performs a task incomparably more complex than solving a set of equations with 100 unkowns. Serious linguists realized this all along, and from the unsucessful — but not necessarily vain, though in some cases premature — attempts we have learnt to understand it even better (Cf. art. 47).

Computer-based *text editors,* of course, have made life easier for people who write in the rich countries. It is such a delight to be able to conceal one's typing errors by using a delete key and one's stylistic slips by simple insertions and replacements, but we owe none of this to CL. In fact, none of the widespread systems deserve the name of editor, since they only help you implement amendments you have independently decided to make. At best, they help you find occurrences of well-defined amendenda — on the lowest level: There is no widely used system which would bring you the participle *hidden* when you have decided to replace the verb *conceal* by *hide,* and with most of them it is up to you to avoid *concealous* for every *hideous* in your text. A little more intelligence in the system would give much better *support for authors.* But then that little extra must include functions based on at least a rudimentary recognition of context, that is of linguistic structure; then, and only then, do we have the beginning of applied CL. A few recent attempts prove the appropriateness of this approach — but industry still seems to be more interested in text reproduction and updating than in text creation.

A special case — in fact, not so special after all — of authoring is *translation.* We have already commented upon the scarcity of attempts to provide good interactive support tools (see also art. 51 and 54). Another distressingly 'mechanical' authoring task which is as yet beyond the mechanized systems is *proofreading.* So-called spelling correction, today based on mere lexical lookup, is a modest beginning. Indeed it sounds very far from sensational to offer a proofreader's aid — but a good one would probably have a greater impact on the world's affairs than an automatic translator. At the far end, of course, proofreading involves every depth of comprehension and erudition. It is easy to prove the infeasibility of an infallible faultfinding machine. But quite a few errors and inconsistencies can be revealed by applying parsing and matching techniques which are classical in CL.

Whereas systems exploiting in an interactive fail-proof environment a small amount of *text comprehension* on the machine's side are conspicuously few and far between, the text-digesting homunculus who comprehends the text so completely that the text can be laid aside and the homunculus interrogated instead have attracted the attention and resources of many. Needless to say such systems can cover only minute subspaces of human knowledge. A problem is that the restraints are sometimes not well recognized: lawyers, to take one example, who believe that it would be an ideal to establish rules for machines to devour and make conclusions from, rather than fomulate necessarily vague texts for humans to actively interpret have understood law and language equally poorly and ignore the creative aspect of reading in general and law interpretation in particular (exempting again some very stable and formalizable microworlds, such as certain special taxation subsystems, where all innovative thinking can be banned over considerable periods).

An interesting kind of machine intelligence is *image comprehension,* offering us the option of drawing maps, pointing at pictures or even making gestures and still being understood via machines. Some of the techniques for analyzing linear text strings are promising for recognizing and in some sense comprehending two-dimensional images or (two-dimensional representations of) three-dimensional objects. Already the questions in what sense one could possible 'comprehend an image' or 'make a summary map' invoke problems — such as topic/focus — which are familiar to linguists. This is an area where one has reason to expect real break-throughs in a decade or so — theoretically and practically.

A major part of a writer's efforts is engaged in a continuous battle against unclarity and ego- or group-centricity. Efforts in the opposite direction are those towards maintaining secrecy. When so much of what we write and say is physically accessible to so many via electronic channels, *encryption* will be a more important part of writing. Construction and breaking of secrecy systems are classics in linguistic computation, and they are likely to receive renewed attention. Whereas governments and resourceful organizations have not failed to escalate this technology, exploiting linguistic results in a manner not always advertized, little has been done for poorman's everyday text protection against, say, his government (Beth/Heß/Wirl 1983).

4.1.2. Reading

As with writing, very little has been done for the primary function of reading (in the sense explained in 4.). If anything, computerization has given us large amounts of poor typo-

graphy and layout, flickering screens and an incredible graphic monotony which no one would have accepted if it had not been backed by the status of computers. Now, this is a passing technological phase, and in any case linguistic sophistication is surely not what is needed to regain the Gutenberg level. But there are recent trends in the positive direction. Some systems use colour to differentiate and highlight text portions of different prominence. It seems that it was British Viewdata that brought colour to the computing world on a larger scale as a compensation for lacking font variety on low-resolution screens, Viewdata being the first really large attempt to offer computer-based reading to a broad public.

This and other similar systems also add movement — oscillation, flashing etc., — to its presentation. There is reason to believe that *text enhancement* will be a rapidly developing area. In addition to or in lieu of conventional typographical means so much can be done, once the text is housed in a computer to help the reader. Since empty space does not cost, there is no reason to display text as compact pages but sentences can be presented in structured ways — those familiar with LISP programming will think of PRETTY PRINT — with subclauses graphically separated or, like footnotes, concealed until called for, with important words or phrases highlighted, with syntactic and other structures clarified by position, font, size and colour of the constituents, etc. etc. These ideas are not new. We have examples from runic stones of colouring according to word classes, and journalists designing front pages of a good daily newspaper know much about this matter. But the resources in a computational environment offer a real challenge.

If little has been done to help reading as such, much has been done to help us not-read. That is, incidentally, what *retrieval* systems can do for us. We need no particular technology to read the lot; the retrieval function must not be confused with acquisition. Problems of selection lead deep into linguistics, to understand the query, to analyze the retrievands and to make appropriate matching. The 'qualitative approximation' involved is essentially a linguistic issue, which cannot be replaced by implication and other operations on 'exact' meaning representations. (Retrieval problems are treated in art. 55—58; cf. "Homeosemy" in Walker/Karlgren/Kay 1977.)

An interesting kind of aid to readers are *text generation* systems, producing some kind of text from non-text materials, such as tables, formulas or even maps and images. It would have helped when such systems are being designed, could we have specified what is assumed to be a coherent text. As a first approximation it is thought to be a linear sequence of symbols which makes sense when read from left to right — except that full linearity may be exactly one of those tiresome features of many man-made texts which we want to eliminate — a feature which is at most fictitious in, a law book or a computer software manual. This is a hot area where really new results should be expected (cf. art. 37 and 48).

4.1.3. Interacting

The more we study them, the more it appears that the difference between reading and writing is exaggerated. Many activities entail *reading and writing.* A writer needs support for elaborate reading functions — to retrieve other texts or passages of the text under work; to consult dictionaries, term banks or precedence cases if he is a translator; to compare with law reports or model contracts if he is a lawyer, etc. — and, convesely, a reader needs to tell his dialogue partner about himself, his needs and background and preferences, for his queries to be properly handled. A special case, treated in art. 42, really defeats our ability to tell whether it is writing or reading: text reconstruction.

Recent linguistics has shifted focus from transmission of messages from sender to receiver to *dialogues* as a kind of interaction between two parties, in turn provoking each other to new responses (Kiefer 1983). Even where the aim of the interaction is unidirectional information flow, e. g., where a client is consulting a librarian or other expert, both are asking and both are answering, building up a basis for understanding. In fact, every single utterance can be seen as a question as well as an answer, blurring even this fundamental dichotomy. Linguists begin to have something to say on *pragmatics,* i. e., on how a communication process should be organised to make the components interact properly, taking terms, handling circumstantial information which has little to do with the underlying subject matters and yet needed for the traffic (cf. Ungeheuer 1972, 17—18). A generalization with new sets of hard issues are the *polylogues* where many talk to many: adver-

tising their interest to learn about a topic or offering information, presenting themselves as individuals or groups in a crowded and noisy market place.

Again the perspective shifts: from two-party interaction between man and machine to man-to-man *communication via machines.* Computing systems assume the role of storing and distributing documents (messages of any size), like newspapers, library and second-hand bookshops all in one, joined together to worldwide communication networks. Techniques, based on what we know about human dialogue and polylogue, will have to be designed for entering queries or submitting cues for others to find our offers in such an environment, where an intelligent machinery connects people at different sites, with different perspectives and manners of expression — and at different times: questions may precede answers in time or vice versa. Particular attention will have to be given to restraints for the information flows: how can unauthorized transfers be inhibited, and who shall have the authority to inhibit what? (Karlgren/Walker 1983)

In an advanced computational environment interactive reading/writing techniques may dissolve the very concept of document: if the reader receives information selected and shaped in tune with the description of his needs, and the author, accordingly, submitted not a coherent ready-made text but rather a response pattern, a kind of text-generator, the output of which he will never see, then who 'wrote' what actually appears before the reader's eyes? Who is responsible for errors and omissions? If we assume, as we must, that both use software which neither of them wrote or understands, if we further assume, as again we must, that host systems impose their own transformations, restrictions and clarifications, we shall need all our ingenuity and all our knowledge about language not to lose control. How, to take only one urgent task, could we construct a tool which would enable two persons to verify whether the 'messages' one sent to the other have been wilfully censored along the route?

4.2. Designing Tools for Language Learning and Teaching

Any kind of computer-aided instruction (= CAI) requires advanced dialogue techniques, which in turn implies linguistic techniques regardless of the subject-matter. Teaching machines were in vogue some thirty years ago, and left the scene with a rather bad reputation: most of what they could do could be done equally well or better by books, given that the same amount of preparation was invested in the tool, and they encouraged fact-oriented drill exercises which were deemed at the time to be incompatible with personal development. However, CAI in general and computer-assisted language instruction in particular are coming back. It seems that pedagogy has become slightly less dogmatic; and computer-based instructors need no longer be rigidly tied to a few correct/false alternatives nor at all confined to grammar rules and vocabulary. Interesting developments are computerized text evaluation in interactive authoring systems, and language training — including of course training in spoken communication, preferably telephone conversation — connected to communication systems, encouraging group dynamics and activating backbenchers. The applicable linguistics, then, is not only parsing and sentence generation but also pragmatics. See further in the contribution on CAI in art. 60.

It should be remembered that teaching of artificial languages has been computer-aided ever since interactive systems became available at reasonable costs. In fact, there is an ingredient of teaching in every interactive system, though often otherwise labled, e. g., in so-called expert systems marketed to resourceful adults. One could even say that instruction to users is a necessary feature of any dialogue worthy of the name.

4.3. Designing Languages

Although a linguist learns to respect the received languages more the more he learns about them, he need not want to leave the world of languages as he found it. What greater challenge could there be for him than to sift out the essential of what we know about human linguistic behaviour and construct a tool appropriate for a given purpose? The whole spectrum of CL insights is relevant for the task of designing a language for a given task be it called programming language, query language, command language or something else.

As already stated, there are deliberately designed features in natural languages as well. Dictionaries and term banks are instruments to control vocabulary, and the methods of CL are used widely in lexicography today, as described in articles 38 and 39. Computational methods have also been de-

veloped for evaluating term-inventaries and even suggesting new terms which fit a given language. We can mention here similarity examination of proposed new trade mark words (to avoid possible clashes with confusingly similar though not homonymous earlier marks, Brodda/Karlgren 1969) and computer-aided coining of new marks or company names which satisfy given general specifications. In Sweden, even new family namens have been computer-generated en masse according to established patterns, to replace the over-frequent namens ending in *-son.* A book, listing computer-generated available names, guaranteed to be new but conventional-sounding, is published by the Patent Office since the 1960's (Molde 1964).

Language planning — or language maintenance, 'Sprachpflege' — is restrained by the planners' incomplete knowledge about usage and the users' incomplete knowledge about the decisions made on their behalf, by linguistic ignorance which causes the attention to focus myopically on petty details, and by the public's reluctance to take directions and the planners' diffidence as to their real power. All these three obstacles are shrinking: in a wired society, where the majority of written and many spoken messages are formulated, revised, transmitted and stored in interconnected, electronic networks, language can be mapped with precision, and norms (complex linguistic rules as well as wordlists) can be made part and parcel of text editors issuing recommendations or warnings and, when nothing else helps, inhibiting transmission of improprieties; better insights in the mechanisms of language and the interdependence of language components will make control action more meaningful and effective; the hesitance to issue and the reluctance to obey orders seems to be faltering in so many places. CL will find large new markets in language engineering.

4.4. Observing Language Use and Language Users

Observing language behaviour is, of course, a major task in linguistics and in some other social sciences. CL is here applied, i. e., it serves as an auxiliary in pursuits other than its own (see, e. g., the articles 40—46). Particularly prominent is the support of the work of the lexicographer, no matter whether his goals are practical or scientific (see articles 38 and 39).

Observation of usage is also a necessary part of the work to design and redesign languages, on which we have already commented above under 4.3. In information science, 'user studies' have become an accepted branch. But on the whole, more effort has been spent on advocating solutions than of their empiric *evaluation.* Self-recording interactive environment will make this attitude less defensible.

A primary linguistic task is to *recognize* a speaker or writer. Our ability to do so — one of our primary capabilities as a biological species — has proven surprisingly hard to parallel in algorithmic machinery. Some of the earlier claims about voiceprint (article 47) and about authorship determination (article 41) can be rather easily falsified.

Applications are not only forensic and philological. It would be safer and so much more pleasant to abolish passwords and directly address electronic assistants, librarians, bankers, etc. who recognize their partners personally and remember how they prefer to be met. In fact, advanced dialogue systems will have to include some personal recognition features. No one will become a good conversationalist unless he learns to recognize faces and voices; this is true for machines as well as for humans.

Going a few steps further, a true dialogue system will have to develop the ability to recognize character: Is he/she a diffident beginner or an oldtimer bored by redundant explanation? Is he/she meticulous in typing or should he/she be asked to confirm? Is this a person who would appreciate examples more than a concise rule? These questions are inevitable in an adaptive and cooperative system, whether or not it is openly pedagogic or not (Kaplan 1977).

Observation need not be restricted to one dialogue partner at a time nor to details about language, knowledge and preferences. With a profession's or a whole nation's written and oral interaction occurring in a self-recording electronic environment, the road is open to constant surveillance of group dynamics, applying all what is known about linguistic pragmatics, content analysis and all other arts of the métier to record: Who is writing, talking to whom in what tone? Who knows — has ever heard, has heard more than twice — this word? Who has read what? Who is aware of what? Who is alerted by what kind of cue? Who has a negative attitude towards such and such words? — When tools for linguistic analysis become less crude and storage capac-

ities virtually unlimited, it will be possible to map every individual and his relationships to individuals, words, concepts and attitudes in greater detail than he could possibly know himself, to predict his next word or response with a precision which would amaze him would the system reveal it to him. What economy could be achieved on such a solid foundation of knowledge and meta-knowledge! How appropriate answers every one would get to all questions! How many unnecessary messages eliminated! What superb social control and coordination could be attianed!

5. Literature (selected)

Aarsleff 1975 · Allen/Perrault 1980 · Anttila 1972 · Arndt 1967 · Bátori 1977a · Beth/Heß/Wirl 1983 · Brodda/Karlgren 1969 · Bühler 1934 · Cassirer 1923—31 · Cherry 1957 · Codd 1974 · Donato 1980 · Dutens (ed.) 1768 · Enqvist 1987 · Fraser 1894 · Greenberg 1966 · House 1970 b · W. Humboldt 1836 · Kaplan 1977 · Karlgren 1970 · Karlgren 1973 · Karlgren 1979 · Karlgren 1981 · Karlgren 1986 · Karlgren/Walker 1980 · Karlgren/Walker 1983 · Kay 1980b · Kiefer 1983 · Koerner 1973 · Lakoff 1972 · Lenders 1976 · Lenneberg 1966 · Lieb 1980 · Lieb 1983 · Loemker (ed.) 1969 · Meyer-Eppler 1969 · Paul 1909 · Pinkal 1983 · P. Schröder/Steger 1981 · Shannon 1949 · Shannon/Weaver 1948 · Ungeheuer 1972 · Ungeheuer 1974 · Ungeheuer 1977 · Verburg 1951 · Walker/Karlgren/Kay 1977 · Whorf 1956 · Wunderlich 1974 · Wyld 1932 · Zadeh 1975 · Zadeh 1982.

Hans Karlgren, Stockholm (Sweden)