Past and future goals of Computational Linguistics

Winfried Lenders

University of Bonn, Germany
Institute for Communications Research and Phonetics

e-mail: Lenders@uni-bonn.de

0. Abstract

For different reasons, Computational Linguistics is at the beginning of a new era. One of these reasons is that the technological progress makes it possible to process complex algorithms in real time and to handle large amounts of data in a such way that storage size and fast data access are no longer the bottleneck of language processing. The technological progress makes it possible to recognize the regularities of language production and language reception from the statistical probability of their occurences.

A second reason is that Computational Linguistics is no longer restricted only to the channel of orthographical language. The restrictions of this channel have dominated research for more than 40 years. We now are able to use parameters of spoken language like word accent

and intonation patterns for the description and disambiguation of sentence and discourse structures. In the near future the visual channel will be available for transcribing gestures and face expressions and to use them for structural analysis.

From a third point of view Computational Linguistics becomes more and more interesting for commercially applications. It is not only the old idea to overcome the language barriers through mechanical translation, but also the growing needs of our society, to get information as fast and precise as possible, which require new products to decode text and speech (e.g., speech recognition, speech synthesis and information extraction).

On this general background this paper discusses the position of Computational Linguistics between academic research and technological applications. On the basis of a detailed analysis of the history and current situation of Computational Linguistics future goals and trends of academic research are to be outlined. Especially the progress of computer oriented discourse analysis and of multimodal research methods will be mentioned.

1. Basics: Computational Linguistics between research and technology

Last year a new "Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" by Daniel Jurafsky and James H. Martin appeared. In the foreword of this fundamental book the authors declare, that "computational linguistics has a forty-year-history as a part of computer science" and that "it is only in the last five years that language understanding has emerged as an industry reaching millions of people, with information retrieval and machine translation available on the internet, and speech recognition becoming popular on desktop computers."

Although in this quotation the industrial and economical point of view seems to be the first point of interest, the authors are aware that "this industry has been enabled by theoretical advances in the representation and processing of language information" (p. xx). Computational Linguistics, indeed, above all is a theoretical field, which is concerned with language and communicative behavior. Another author of a recently appeared introduction to Computational Linguistics, Roland Hausser, declares: "The goal of computational linguistics is to reproduce the natural transmission of information by modeling the speaker's production and the hearer's interpretation on a suitable type of computer. This amounts to the construction of autonomous cognitive machines (robots) which can communicate freely in natural language." (Hausser 1999, p. 1)

Following these quotations Computational Linguistics primarily is a field of academic research and secondly a field of technological systems construction. But there is no doubt that the fundamental impulses came from outside of research and that the results were always evaluated with respect to the central application fields. In addition, in recent years there is an increasing competition between basic research and technology, and therefor it makes sense to reflect what are the progresses which we made since the beginning of our field, where we are standing today and what are the goals for the near future.

2. Application fields

The history of Computational Linguistics may be characterized as a network of mutual dominance of certain application fields on the one side and maintream linguistic topics on the other. In the following in chapters 2.1 und 2.2 I intend to characterize these corresponding lines in Computational Linguistics from my point of view. Later on, in chapter 3, I'm going to suggest some fields of theoretical and practical interest which are currently under consideration or still missing in Computational Linguistics.

2.1 The paradigm of Machine Translation and the desire for overcoming the language barriers

2.1.1 The technological situation

When visiting the Alta Vista Homepage you can find among other links one which leads to a translation tool called BABELFISH. This tool enables the user to translate a particular webpage from a source to a target language, e.g. from English to German. Some of you may know that the name of this internet translation system, Babelfish, is taken from one of the most famous modern science fiction novels, namely from Douglas Adams: The Hitchhiker‘s guide to the Galaxy (in German: „Per Anhalter durch die Galaxis“), first published in 1979 . In this novel Babelfish is something like an ear plug which a person can put into his/her ear and which is able to translate all languages which the hichhiker is confronted with. Babelfish, in our science fiction story, realises a dream or an utopia which is as old as the history of man. The non-existence of a common language for universal communication at any time was felt as a big and inconvenient obstacle in communicative contact between people and nations. In the European mythological tradition the Babylonian Confusion of Language as a consequence of mens arogance against god is a well known symbol which is combined with the story of the tower of Babylon.

Considering how to overcome the language barriers was always a fascinosum of mankind. And it is this fascinosum which has inspired the grandfathers of machine translation 50 years ago and which finally led to the foundation of Computational Linguistics als an academic discipline.

After 50 years of research we are now in a situation that machine translation is no longer a main field of academic research, but of technological and industrial development. Babelfish ist not the only translation system in the world. John Hutchins, perhaps the most important expert concerning the history of machine translation and its current situation has found more than 290 systems which are available in the world as products and which are produced by more than 180 companies (Hutchins 2000). This number involves just the products and not the experimental machine translation systems and translation aids, and most of these products are determined for use in the internet. Some of the firms where you can buy translation systems are specialized in machine translation like SYSTRAN, others are affiliations of larger companies like LINGUA-TEC, and a third group offers machine translation tools together with other language processing programs like speech recorgnition, as Lernout & Houspie Speech Products.

It is overwhelming, what language pairs are available. Of course, most of the systems are with English as source- or target language, but there are also existing systems for French, German, Spanish, Italian, Japanese, Arabic, Chinese, Finnish, Hebrew, Hungarian, Korean, Russian and Vietnamese.

In addition to those systems which we can use in the Internet, we have available translation programms in every bookstore, cheap and easy to install on our PCs. There seems to be a market for these systems, and there are a lot of evaluations and tests in computer journals and newspapers. Observing the development of the last 5 years, you can recognize a real jump in translation quality, robustness and speed, but, and this is the crucial point which makes MT at least suspect, the quality for many purposes obviously is not good enough and systems fail or break down at most simple linguistic problems. After 50 years of research there is no getting round to notice that machine translation is far from what human translators are able to perform.

With respect to this situation, from an academic point of view it is necessary to look were we are standing today in this field of research.

2.1.2 Short overview of the history of machine translation

Although the idea to construct a mechanical device for the translation from one language to another occured several times during the history of mankind, realistic machine translation systems came into consideration since the end of world war second. The origin of machine translation was the understanding of language as a computable phenomenon, a conception, which came from the American information science (Shannon, Wheaver) who thought that translation could be done by statistical methods and by methods from information theory..

Form our point of view the history of machine translation from these beginnings can be roughly devided into three phases.

The first phase from 1945 to 1954 is often identified as the era of Word-by-word-translation, ending in 1954 with the Georgetown-IBM-Experiment (Russian-Englisch), where a selection of 49 sentences in restricted language and with restricted dictionary and grammar was presen-ted. The Georgetown-experiment as a whole was a presentation of 4 single projects, one of them became the later SYSTRAN-System, which can be used in the Internet under the already mentioned name "Babelfish".

The second phase from 1954 to 1967 is characterized by the introduction of syntactical analysis. As a result of the Georgetown-Experiment and of its optimistic perspectives during this phase many machine translation projects were founded in the United States, using the newly developed formal grammars.

The third phase began after the ALPAC-Report from 1966/67 (Automatic Language Processing Advisory Committee) which summarized and evaluated the results of so far MT systems. The main results of this evaluation were that a high quality fully automatic translation (HQFAT) seems not to be possible and that machine translation demands much more basic research in linguistics. Although as a consequences of the ALPAC-Report most of the MT-projects in the US were canceled, basic research in Computational Linguistics increased, in particular in the semantic level and with respect to the problem of understanding.

A fourth phase, from 1974 to 1995, is characterized by the European and Japanese initiatives in machine translation, and, at the end of the 1980-decade, the integration of spoken language. The main motivation of this phase came from the increasing worldwide traffic, trade and communication and from the new insights into the nature of human language unterstanding processes.

From 1995 in my opinion we may identify a fifth phase, when the first PC-Versions of Machine Translation systems and Internet systems appeared and a broad commercialization of Machine Translation began. This phase may also be characterized as the beginning of a more technological approach to machine translation.

2.1.3 Goals for research in Machine Translation

Machine translation as we have it available now in many forms and products is the result of 50 years of academic research. But today we have only a very few research projects in this field, because MT became - more or less - a field of technological construction. There are several reasons which make this plausible:

A first point is that serious researchers are aware since several decades that a high quality machine translation without human pre- and/or postediting is impossible in principle. They know that, in order to make MT systems more perfect, further research is necessary to prepare better lexical and grammar resources. Therefor, as a consequence of this insight, we now have worldwide a lot of groups working on different aspects of language technology like automatic tagging, morphological analysis, grammatical tools, etc. And also empirical linguistics, based on corpora and corpus description came back into favor.

Secondly, increasing processor speed and memory capacities allow very fast dictionary lookups and complex parsing processes. At the same time computer systems became small and popular, therefor low cost MT-systems became realistic and got a chance on the market. The translation quality of these systems in general is not sufficient, but the fascinating idea of MT promotes the marketing of these systems. It is characteristic for the current situation that there are also some high end systems, which offer complete dictionary services. That means that dictionaries for special languages, terminological data bases and translation memories are produced by these firms, including the necessary support. Systems are optimized and tuned according to the respective needs of the customers, and, of course, this is only available for a much higher price. Fast processing and high memory capacities allow the introduction of large translation memory systems (TM) which make machine translation much more effective, applicable and robust. From a linguistic point of view, however, a translation memory is a simplification of the translation process, since a translation memory is a computer storage in which already conducted translations can be stored and reused for further translations. It works like a dictionary of sentences or phrases, and must be embedded into an appropriate text processing environment. TM-systems are in fact machine aided translation tools, and they usually allow the intervention of a human being. They contain not only idioms in a traditional sense but phrases which occur frequently in a technical environment and sentences which are acquired from aligned texts. Translation Memory techniques do not have to respect the natural linguistic boundaries. They just stock paragraphs, sentences, or fragments of sentences under the only condition is that these correspond to a certain level of repetition. This level of repetition can be identified by statistical and structural methods.

Third, providers of translation software shift the responsibility for good translations to their customers: They instruct them to do a pre- and postediting by reducing the vocabulary and the syntax of their texts. Regular syntax is substituted by a reduced syntax which may allow e.g. only main clauses, no relative clauses, no complex sentence patterns. The author of a text is instructed to formulate more simply and more clearly, i.e. to formulate his text with respect to the expected automatic translation.

A fourth point is that, although we had several knowledge based translation projects, the currently available translation systems have no idea about world knowledge, knowledge about the situation or knowledge about the communication partner. We know that this type of knowledge is necessary for the understanding process, but up to now we do not know how to integrate them into translation systems. It is this fact why there are sceptical people who estimate that mechanical language translation systems "are just not worth the effort", to construct and care them.

A fifth characteristic element of current situation in MT is the introduction of statistically based systems. While 'classical' translation systems intend to explain and to simulate the process of human translation, statistical translation focusses more on the result of a translation process. (Jurafsky/Martin p. 819). Statistical translation uses learning algorithm in order to learn from existing translations in multilingual corpora (aligned texts) the probabilities of equivalence between sentences of two languages. These techniques came into Computational Linguistics when it began to merge with the fied of speech processing, about 10 years ago.

This leads to a sixth point: At the beginning of the 1990 it became apparent that speech recognition and speech production technology would be applicable also in MT. In Japan, in the US and in Germany industrial firms and governments finanzed large projects to construct translation systems for spoken language. These systems should support the spoken dialog of people with different language skills in a restricted domain. The German Verbmobil-project, e.g., intends that, for instance, a German applicant for a conference in Japan should be able to make his conference application by telephone in English, using Verbmobil as a translation aid in the case of German to English translation problems. The Japanese part in this example should also be able to communicate in English with the help of a computer tool in cases of translation problems. That means that the aim of Verbmobil is to support a spoken language communication between a German and a Japanese participant with English as a mediation language. The main problems of these projects are not in linguistics, but in the field of spoken language recognition and generation. Since the intended domain of application, the vocabulary and syntax, so far is relativly restricted, the linguistic problems are not so crucial. The German Verbmobil project was finished in 2000, with many results particularly in the application of statistical methods to MT.

Seventh: What has been said with respect to machine translation concerning commercialization of system, is also true for the field of the speech processing. During the last five years, international software companies detected the possibilities of mechanical speech recognition and speech synthesis. Some of the actually discussed applications are voice-based calling center systems, voice-portals and voice-browsers, which are developed by several firms and which are products in the near future. These systems use the so called say-anything-technology; that means that they react no longer only to simple commands but understand complete naturally spoken sentences. But we are still at the beginning of this evolution. Although there may be some significant progress in the near future, we neeed more fundamental research in natural language communication processes, whether for the written or for the spoken mode.

The conclusion from this analysis of the history of MT and the state of the art is that we are now in the situation, that some linguistic problems like morphological analysis and parsing are resolved to some extent. But there remain many unresolved problems like the structural description of complex sentences, the disambiguation by means of world knowledge, the general problem of understanding and the recognition of inter-sentence-relations like anaphora and discourse coherence. That means, that machine translation seems to be attainable in the field of special and restricted languages. The remaining numerous ambiguities, readings etc. can be resolved only from the context in which a sentence is embedded, or even from the situation in which a text is integrated. Every translation which requires an interpretation of a text up to now is possible only by human intervention.

2.2 The paradigm of Natural Language Understanding (NLU)

The second 'historical' paradigm of Computational Linguistics are Natural Language Understanding Systems. As in Machine Translation, in NLU we had in the past highly important experimental systems and theories about human lanugage understanding processes. But the results, with respect to practically working systems, were relatively poor. Nevertheless, NLU was important for Computational Linguistics, Artificial Intelligence and other cognitive sciences. Natural Language Understanding overcame the semantic barrier of the mid-sixties, and made obvious the enormous importance of knowledge, memory, inference and representation. It is often forgotten, that still in 1967 at a conference at Las Vegas on "Computational Semantics" famous linguists and computational linguists like Yohoshua Bar-Hillel mentioned that it never would be possible to compute the semantic level of language. One year later, 1968, the book "Semantic information processing", edited by Marvin Minsky, came out and gave us a first impression how semantic knowledge could be handled in a computer's memory. In the following twenty years language understanding systems were booming, and we had a couple of working but experimental systems like SHRDLU, LUNAR, HAM-RPM, LILOG etc. in different fields and with different aims of research. We also had one or two commercial systems using techniques of NLU, but these were only marginal in the history of Computational Linguistics.

In the moment this field of Computational Linguistics is connected with the idea to construct what is called "communicating agents" or "intelligent agents". In general agents or agent systems are independendly working programs which are able, for instance, to find particular information in the internet, and to communicate with other agents and with users on the reliability of the extracted information (Dignum/Greaves 2000). There is no doubt that the concept of communicating agents is a metapher as machine translation is a metapher. Since in these systems most of the information is in natural language and since the communicative interaction (at least with the user) is in natural language, these agents need some communicative competence, e.g. knowledge, grammars, semantics, spech acts, speech recognition etc., and this, again, is an impotant future field of CL, where basic research is needed.

2.3 Mainstream Topics of Computational Linguistics

While the inspiration for the computational approach to language and communication came from application fields like MT and NLU, the daily work had to proceed with morphological, syntactic and semantic analysis, grammars, dictionaries and so on.

2.3.1 The pradigm of morphological and syntactical analysis

In any MT-system and NLU-system a morphological and syntactical analysis is necessary in order to find out the constituting entities and the structures of the intended texts. That is why the main activity of Computational Linguistics in the past was in the field of constructing theories for morphological and syntactical analysis and to build the corresponding algorithms. Today we have for many western languages morphological and syntactical parser; we have functioning algorithms, and in principle we can say that the morphology and syntax problems are resolved (concerning the state of the art of morphologigal analysis for European languages see Hausser (ed.) 1998). These algorithms are able to identify structures and entities with relatively high precision. The morphological analyzers are lexicon-based and use finite state transducers, or analysis-by-synthesis programs. The syntactic parsers use sophisticated grammar formalisms, which are constantly optimized, or they detect syntactic regularities by using statistic methods.

Problems which not have been solved up to now are under consideration: there is a persisting research which tries to find solutions also for the currently not yet resolved problems like anaphora, long distance dependencies and complex subordinated clauses.

But of course, there are limits: it is not yet possible to disambiguate sentence structure when the disambiguation is only possible from world knowldge and/or situational context. These are, again, fields of current and future study.

2.3.2 The semantic turn: lexical knowledge and world knowledge

Concerning the semantic level, it was already mentioned, that for a long time, nearly 20 years, it was not clear whether it would ever be possible to represent the meaning of words and sentences in a formal way. At a conference, 1968 at Las Vegas, with the titel "computational semantics", famous representatives of the field declare that semantic processing is not in the scope of CL, and that decades of research are needed in this dimension. But it was only one more year, 1968, when Marvin Minsky and his group showed how to handle the semantic dimension, how to represent meaning (by semantic relations) and how to infer from knowldge bases. And it was only a couple of years more that the concept of procedural semantics was established and integrated into experimenal understanding systems (by Woods and Winograd).

The main question, how to represent the meaning of words and sentences were answered: Meaning has to be represented by semantic networks or other predicate calculus based representation language, and the process of understanding a sentence consists in the mapping of that sentence to an already existing representation of meaning and knowledge. As a consequence of the 'semantic turn' in the seventies new theories about the structure of human memory came out and especially semantic nets and other representation languages enjoyed more generally attention.

With respect to the computation of language, representation languages had to be developed as well as large lexical knowledge bases from which people thought that they could be used for any application (e.g. the Japanese EDR-System). This led to another topic which for several years was a mainstream topic in CL: studies on the reusability of lexical information.

The idea was to use already existing lexicographical data in order to construct lexical knowledge bases. This idea was motivated by the fact that more and more printed dictionaries were available in machine readable form and that it should be easy to parse their lexical entries and to export their lexical information into a database. As a consequence from this idea all over the world projects were conducted to exploit the existing material; these projects are called "reuseability studies". The resulting lexical data bases should be theory independent and usable for several purposes. After one decade of research it became clear that existing dictionaries only in a limited sense can be used for automatic extraction of information.

The state of the art in this filed is that there are some international efforts to build up an universal lexical knowledge base, called WordNet (Fellbaum 1998), in particular EuroWordNet. WordNet/EuroWordNet (http://www.let.uva.nl/ewn/) goes back to the original idea of semantic memory, and tries to describe the meaning of words by means of semantic relations. It is realized as a multilingual Network which contains the basic concepts of eight European Language (English, Dutch, Spanish, Italian, German, French, Estnic and Tschec). The language specific concepts are interrelated by a language-independent interlingual record, and the applications are disambiguation, information extraction, lexicography and Textklassifikation.

3. The Future of Computational Linguistics

What are, on this background, the future perspectives of Computational Linguistic research? Let us consider them by first outlining the present technological situatione and then by sketching some promint lines of academic research.

3.1 Technological conditions of CL-Research in the age of multimedia

The situation of Computational Linguistics was always depending on the state of the art in computer technology. Since 5 - 10 years, we are again at a turningpoint, which is closely associatied with the development of multimedial und multimodal computersystems. Computational Linguistics for several decades got its decisive impulses from the utopias of mechanical language translation and natural language understanding systems, but in most research projets up to now it was limited to written texts. In future, thanks to the rapid technological development, we will have at our disposal for research also the acustical and the visual channel of communication. That means that we are able to study the role which is played by intonation as well as non-verbal forms of communication like gestures and postures. This opens the chance to get more criteria from real language behavior in order to resolve ambiguities and to find the correct entities and structures of our utterances.

3.2 New application fields and visions

What are the most important application fields? Is there a vision which may serve as a stimulating metapher for our intended models?

It is not more than a platitude that we are currently living in a social environment in which the manipulation and managing of knowledge becomes more and more important. This was true 30 years ago and this is true in particular for the present internet society and for the future. The human society with its global exchange of information and commercial products depends on fast and powerful tools for managing of information and knowledge, and this most probably will increase in the future. From this importance of knowledge and information management the idea of mechanical or automatic handling of knowledge was born. Since knowledge is mostly encoded in natural language, this, of course, from the beginning, was a main topic of Computational Linguistics. So, 25 years ago we had many projects in the world dealing with automatic abstracting, indexing and retrieval (Sparck Jones/Kay1973), and at present new efforts are made to construct new systems for information extraction and knowledge management. In some respect we have better conditions: we have the international standardization of information using standardized markup languages like SGML and XML, and we have tools for morphological and syntactical analysis. What we not yet have and what is highly important for information extration are fully developed tools for intersentence analysis, discours analysis.

Another important point in this context are storage and access to those information which were traditionally and for hundreds of years encoded in texts, dictionaries and lexica. How shall they be managed in the future? How can we use modern technology to manage the historical memory of our human society?

Modern text and database technology in conncection with multimedial and multimodal representation enables new concepts for structuring and organization these types of information or knowledge. Technology offers new opportunities to organize and present the data, which in turn offers expanded possibilities for the user to function more efficiently and more effectively. For instance, textual resources of the past, manuscripts, images, sounds etc. can be integrated into a hypertextual environment which opens the possibility of a significant increase in the quality of the information and which represents the beginning of a new century of textual studies.

The idea to have systems which allow to extract and to manage textual based knowledge has to be connected with a second idea, namely to do this not only with written but also with spoken texts. An in this respect a new utopia comes into vision, which from its origin is a philosophical one, namely the vision of a future oral society. There are some media theories, propagated by famous people like Marshal McLuhan and Villem Flusser, which came to the result that the global society of the future does not need any written form of communication. From this vision a new model and new tasks for Computational Linguistics become visible: the automatic transformation between different communication medias, e.g. the transformation between speech and writing, or more precisely, the managing of knowlege in spoken form without any written codification. This vision of a future society may be in a similar way fruitful for Computational Linguistics as it was the utopia of machine translation for the 50-year evolution of our field. In my opinion in order to establish the utopia of a writingless society we have to find new methods to process communication, i.e. to store the contents of our communicative activities, to accumulate and to distribute and to circulate human knowledge. This is the challenge of today for Computational Linguistics. It is just a vision, but as a challange it is in principle quite comparable with the utopia of mechanical language translation.

3.3 About statistics

In recent years we sometimes had a controversy between rule based systems and statistically based systems. In my opinion such a controvery does not recognize the fundamentel interdependence of what can be seen from statistical regularities and what is the reason or the basic fundament of these regularities. Both aspects, the rule based aspect and the statically based aspect reflect one basic property of language and communication by different methods, namely that communicating people use rules and rule systems for their communication and do not communicate randomly and arbitrarily. As for all phenomena, also for language it is true, that "statistical methods are only approximations to the underlying data generating mechanism." Statistical methods are useful to help in understanding what is going on. (Lindsay 1995, p. 97; from Wolters 2001). That means for language studies, that statistically based systems indicate that there are underlying rules which could be detected and formulated as such.

To find these basic regularities of communication and to concrete them into a theory is the overall goal of linguistics and communication theory. The goal of Computational Linguistics is to formalize these rules in such a way that they can be programmed and run on a computer. Running on a computer means simulation in a certain sense, and this is a particular method of computational aided research. It has often been used as a means for optimizing a formal system or theory. This in mind, it is irrelevant whether you find the theory by statistical methods or, e.g. by minimal pair analysis. The goal is the theory, and the goal are the rules. Statistics reflect rules, they may be obvious or not.

Working systems which should be applied or used outside of science and research are necessary byproducts of research. For instance, a machine translation system is a byproduct of translation theory, morphological, syntactical and semantic theories, etc., and in a translation program these theories are combined with technological know how, i.e. know how to organize rapid access to big computer memories, fast dictionary lookup, optimal algorithms and robust solutions.

From this point of view statistically based systems are primarily located in the field of language and speech technology, which is strongly connected with Computational Linguistics, while the rule based systems are more or less the domain of basic research.

3.4 Future research fields of Computational Linguistics

From the above mentioned state of the art of MT, NLU and others, we have seen that the remaining problems are still those of ambiguities: ambigues words, sentences, relations between sentences etc. If we don't want to resolve these ambiguities just by probability scores, we have to do extensive further empirical studies in order to find the underlying rules of communication and understanding. That means first that we have to study much more of the linguistic behavior than just sentences and secondly that we have to study this by the investigation of corpora. To be concrete: we have to investigate discours structures, and we have to do these investigations by using all available modes of information.

3.4.1 Discours studies

What is a discours? A discours generally is a set of more than one interconnected sentences. Isolated sentences which are randomly succeeding one another do not constitute a discours. Turns in dialogs do not automatically interupt such an interconnected sequence, so that either monologs or dialogs may constitute a discourse. The interconnection of sentences in a discours may be the result of a mental combining certain ideas which may be called discours entities (Wolters 2001).

The interconnection of sentences is established on two levels, the level of expressions and the level of coherence. On the level of expressions in a particular discours the referents or discours entities are refered to by referring expressions (e.g. nouns and pronouns). With the level of coherence is meant that a discourse must have an internal structure which guarantees that the discourse is coherent. A coherent discours structure consists of discours entities as interconnected parts and certain discourse relations between them. Coherence is mostly claimed as the central property of a discourse.

Referring in a discourse is often ambigues, and also finding out the entities and relations, which establish coherence, if often not possible just from the text. That is why an additional property of discourse has to be mentioned: A discourse is always integrated in a communicative interaction. That means, that a speaker, a hearer and the discours situation must be taken into account when studying discourses. Human individuals, interacting in a dicours, understand one another by interpreting the different types of behavior coming from the particular situation and from the hearer. For computational studies it is therefore most important to take these situational circumstances into account and to include information about the communicative environment.

As other studies in linguistic structure, discours analysis has to use corpora as well. Special corpora must be developed with regard to the particular requirements of discours analysis. These recources should contain not written information, but spoken texts like monologs, dialogs etc., from different domains, and they should also provide a visual channel with the recorded situation and event. The annotation of these recources should provide not only a part-of-speech- and treebank-tagging, but also information on participants, relative position of the participants, moves, intonation, gestures, eye movement, etc.

According to the normal progress of empirical studies also discours analysis, as a main field of research in CL, runs off in three steps:

- establishing of advanced theories on entities and structural regularities of communicative events,
- construction of multimodal resources as an empirical basis for research
- application and optimization.

3.4.2 Advanced discours theories

Having the resources, which offer simply the raw data on what is going on in a discource situation, the next step is a complex annotation on the level of discourse structure. For this purpose a set of dicourse tags must be explored, and such a tag set must be the result of hypotheses on discours entities and structures. In other words: Annotation of corpora for discours analysis should reflect some theoretical approach to discourse and some elaborations on the basis of theories about discourse structure.

Concerning theories, there is already significant progress. There are a couple of already existing discourse models from the last 20 years (Kamp 1981, Heim 1983, Webber 1978, 1983, Sidner 1979, 1983; Joshi/Kuhn 1979; Joshi/Weinstein 1981, Grosz 1983, 1995), and there are also several tools which have already been used for discourse annotation. A frequently mentioned example is the rhetorical structure theory (RST) by Mann and Thompson, (1987) which was frequently used for practical discourse annotation. RST is a descriptive theory which can be used to describe the structure of a coherent text by means of about 20 sets of rhetorical relations. By annotating these relations to appropriate textual units, a tree structure of the text can be generated. Fig. 2 shows an example. The original theory by Mann and Thompson was adopted by Marcu (1997), who created a program for the annotation of texts with RST relations and who experimented with a measure of importance which should allow to extract automatically the most important units from the discourse structure. The theoretical assumption behind these experiments is that there are relations between the kernels of a coherent text, that these relations can be identified and be described explicitely so that they can be used by a system to

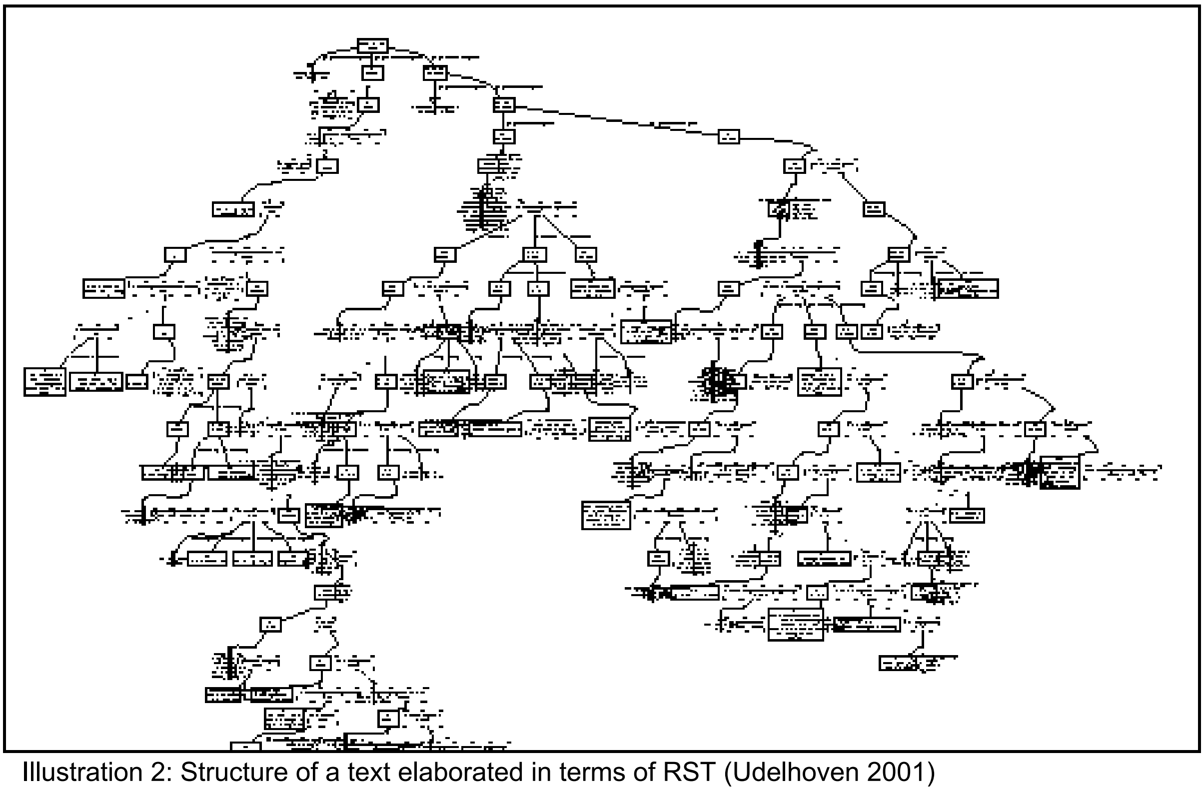


Illustration 2: Structure of a text elaborated in terms of RST (Udelhoven 2001)

extract the relevant information. Of course, this system, when finally constructed, would operate as an intelligent and autonomous cognitive machine with linguistic competence, in the sense of the above mentioned definition by Hausser 1999.

Another option to describe discourse structures is the systematic use of argumentation theory to linguistically annotated discours corpora. As an example, fig. 3 shows a very early experiment with some philosophical argumentations (from Lenders, 1974). In this example a philosophical text has been annotated by part-of-speech markers and a set of argumentative operators which represent relations between sentences.

Theories of this type are still restricted to written texts which have to be annotated by particular markers preserved by the respective theory. They open a wide range of possible studies, i.e. the elaboration of text coherence, and discourse structure and they also open some aspects of practical application, e.g. in the erea of information extraction and text interpretation. But because of their restriction to orthographical texts, several dimensions of language communication don't come into their focus, and some of the ambiguities remain. With multimodal corpora new empirically based theories would become possible. Such theories should enable better chances to identify the entities which serve as referents and referential expressions, and the overall discourse structure. And also the identification of the thematic structure, the argumentative exploration of a certain subject, the parts of an argumentation, claims and warrants etc. would become more obvi-

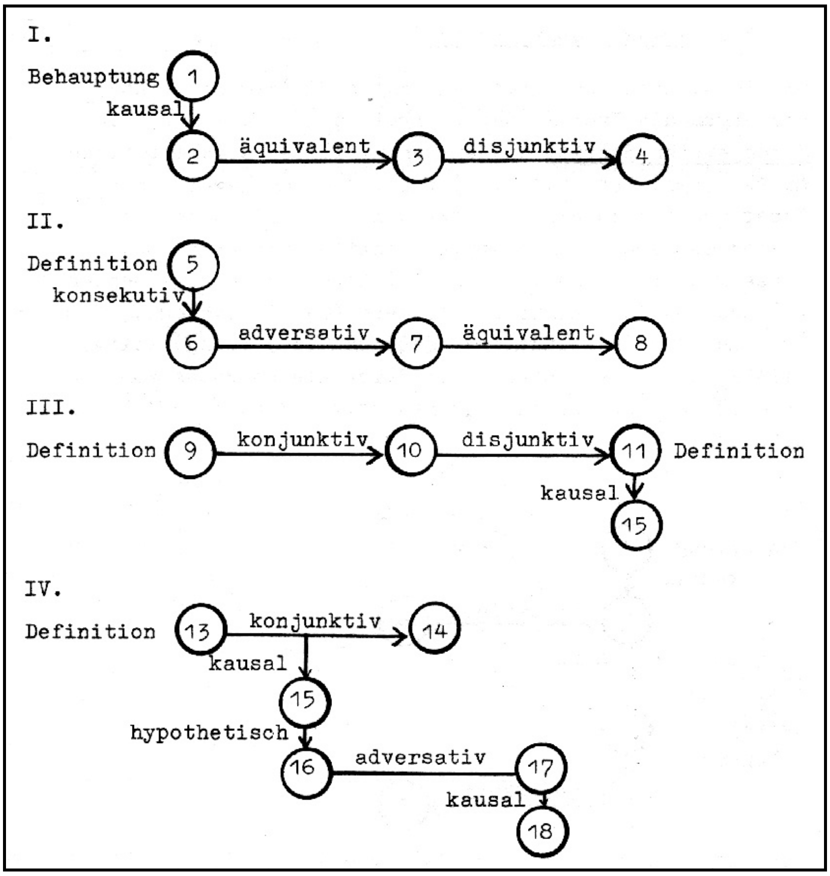


Fig. 3: Argumentation, consiting of four argumentative konstituents (Lenders, 1974, p. 241).

ous. Recources of this type are already under construction (next chapter).

3.4.3 Multimodal resources

To resolve the remaining ambiguities we need more indicators which could not only come from sentence structure but from the overall structure of the communicative discours. In linguistics, first steps to collect corpora of spoken texts and to annotate them systematically were done in the 60th. I remember very well that in Germany several projects were conducted which operated with complex annotation schemas and which were oriented among others into the study of turns and turn taking behavior in spoken dialogs. We also had experiments with video-recorded situations, in order to interpret different nonverbal gestures and to correlate them to the verbal behavior. It is also well known that in experimental psychology we have studies e.g. on the correlation of eye movement and attention. But these studies are out of the scope of CL because they consist more or less of simple audio- or video recordings which have to be transcribed manually.

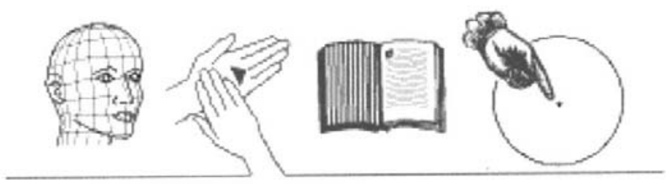
Since several years we now have available systematically established phonetic corpora or speech corpora. There are worldwide a lot of those, and we have our national and international agencies which collect and distribute them - among other corpora.

There are two types, corpora of read speech and corpora of spontaneous speech. Both contain the material in the form of digitalized speech signals. They have to be tagged and annotated according to the particular aim of reseach.

As an example I want to mentioned a phonetic corpus which has been recorded and labeld at my institute in Bonn under the direction of my colleague Wolfgang Hess, the Bonn Prosodic Database (BPD). This corpus contains German read speech and consists of short stories, isolated sentences, and question-answers pairs read by three speakers. The corpus was annotated by three phoneticians whith respect to several research goals. One of these goals was to investigate the role of perceived prosodic prominence which is used by speakers in order to emphasize linguistics units in contrast to their environment (Heuft et al. 200, p. 97). Since prominence as a quantitative parameter of a syllable or a word boundary, syllables and word boundaries had to be annotated by using a scale of prominenc values. In the subsequent study the interdepence of the annotated prominence was related with accustic correlates like prosodic and segmental parameters, and with linguistics parameters. For this purpose the corpus must also contain the respective prosodic and segmental annotations and linguistic annotations like morphology, word class, syntax and focus. The conclusions of this study were that "perceived prominence as a gradual relative parameter can be labeled reliably, has reasonable and distinct correlations to acoustic parameters, can be related to linguistic phenomena, and can be well predicted" (Heuft et al. 2000, p. 113). The last result, by the way, made this study relevant also für speech synthesis, and the results could be used at our institute for text to speech programs.

While this projects is using exclusively a speech corpus, there were also relatively early projects with additional annotation on the situational context and the behavior of the participants.

As an example I want to mentioned the MapTask-Projekt, conducted at the Human Communication Research Centre at the University of Edinburgh and the University of Glasgow since more than one decade (www.hcrc.ed.ac.uk/maptask/maptask-description.html). In 1992, they published so called task-oriented dialogues, on CD, which contained the sound files, dialogue transciptions, and digitized maps. The aim of this project is to "elicit specific examples of linguitisc behavior in controlled conditions" . The basic configuration is that "two speakers sit opposite one another and each has a map which the other cannot see. One speaker the *Instruction Giver* - has a route marked on her map; the other speaker - the *Instruction Follower* - has no route. The speakers are told that their goal is to reproduce the Instruction Giver's route on the Instruction Follower's map." The maps consits of landmarks and there is also a group of particpants who were able to make eye-contact which is recorded in the data. The annotations of the material have been released, and interesting people now can find this corpus annotations represented in XML notations, including dialog structure, part of speech, gaze, landmark references, and eye movements.



Human Communication Research Centre

MapTask is one possible approach to integrate multimodal information There are in the meantime several projects in the world which aim to construct databases with multimodal information or to provide tools for the annotation of these data. For people who are interested in this current research, there is an agency which collects information on multimodal tools and resources, and which has reviewed them, the Danish NIMM-Projekt (Natural Interaction and Multimodal Data Recources) (http://isle.nis.sdu.dk/). This 'agency' has provided information on twelve ongoing projects on "natural interactivity and multimodality aspects adressed by each tool in terms of speech, text, gesture and/or facial expression" (p. 2). As is reported by the agency, some of the reviewed projets are just focussed on speech or speech and text, and "most of the tools addressing gestures have a number of basic functionalities in common including the possibility of viewing the video, adding markup, making synchronization, making time alignment, and extracting information" (p. 2).

With respect to this multitude of projects it is surprizing that one interesting field of research - as far as I know - not yet has been entered by computational methods: the use of existing movies or movie scenes e.g. for discourse analysis. Dialogs in movies are task oriented dialogs under controlled conditions - in the sense of the MapTask-resources, and they are available without problems. What must be done, is an appropriate annotation of dialog structures, gestures, body movements etc. witch has to be developed by theoretical considerations. The technological conditions for these annotations are available, and that is why we are currently planning to start such a project in cooperation with film- and movie experts.

3.5 Conclusion

What are the conclusions from this analysis and perspectives?

From a general point of view we have to ask the question what are the particular research purposes intended with discorse analysis and multimodal resources and what are possible application fields.

As a first goal, multimodal resources bring gestures, facial expressions and other nonverbal behavior at our disposal for discourse analysis. From nonverbal behavior we get additional pointers to discours entities or nonverbal referring expressions. This is important for the elaboration of the thematic structure of texts and for recognizing text coherence. Since these pointers have to be annotated in the material explicitely, and then elaborated systematically an important step is to set up an appropriate tag set or encoding schema, which already envolves basic hypotheses on the entities and the structure of the particular phenomena of discours. The MapTask experiment, e.g., has a set of twelve 'moves' in their coding schema, and a couple of what they call 'conversational games', by which the dialog is organized and which in total could represent the structural elements of a dialog. And this, indeed, are elements of a certain theory about the constituents of human communication.

A second goal is located in semantics. It should be possible to find particular correlations between linguistic expressions and gestures. Communication has it own semantics which goes far beyond the semantic of linguistic expressions and include the whole communicative behavior. During the understanding process we use indicators from the visual channel (as in the case of prosodic prominence from the acoustic channel, see above) in order to identify the meaning of words and sentences or structures of sentences and texts in cases of ambiguity. This is an important goal in linguistics and in communication theory.

The third and last mentioned topic of reseach is in the erea of pragmatics. We know, that we have to observe some pragmatic conditions, under which our communicative acts will be successful. And we always check whether we are successful or not. In most cases we don't do that by verbal expressions, but by using other channels, the acustic channel and some paralinguistic signals, or just the visual channel by nodding our agreement or by greeting. There are some particular facial expressions which accompany the corresponding words; these are e.g. expressions of happiness, depression, anger etc., and it may be possible to recognize from them, whether, what we say with our words, corresponds to our intentional state, - or in terms of speech act theory, whether it meets the condition of sincerity.

This may be enough for the moment concerning possible reseach goals. Finally it should be mentioned that there are numerous applications for discours analysis and theories of discours which integrate the sound and visual channel of communication.

An interesting application field could be the interpretation of sign languages which are used in regular situations and which are used by deaf people. Progress in this field is of similar importance as speech recognition and speech synthesis for the blind.

Another applications is information extraction from spoken text or even from the communicative behavior.

And a third field of applications meets the above mentioned vision of a future oral society. There are already industrial projects which are oriented to support and optimize e.g. the operating instructions for everery day technology. There is a project in Germany, called EMBASSI, finanzed by some industrial firms and by the government which wants to realize some aspects of this idea. The basic idea is to instruct a person not by written text but by guiding him or her through interaction. For instance, if you want to use, let us say, a DVD-recorder, the system tries to find out what exactly you want, by multimodal methods, i.d. by the recognition of speech and gestures. When you point at a certain DVD and utter the sentence "I want to watch this movie" the system is able to recognize and to meet your request. These possibilities, indeed, are good chances for the future!

4. Appendix

A. Links

WordNet: http://www.let.uva.nl/ewn/

Bonn Prosodic Database: http://www.ikp.uni-bonn.de/dt/forsch/phonetik/bpd.html

Other phonetic resources: http://www.phonetik.uni-muenchen.de/Bas/BasHomeeng.html

Multimodal resources: ISLE-Projekt: http://isle.nis.sdu.dk/

Maptask-Projekt:

http://www.hcrc.ed.ac.uk/maptask/ (Maptask-Annotationen)

http://www.ltg.ed.ac.uk/~amyi/maptask/demo_top.html (Demo)

Embassi: http://www.embassi.de/start.html

B. Literature

Bátori, István/Winfried Lenders/Wolfgang Putschke, eds. (1989): Computational Linguistics. An International Handbook on Computer Oriented Language Research and Applications. Berlin: de Gruyter.

Boggs, Roy/Kurt Gärtner/Winfried Lenders (2001): Der arme Heinrich in Multimedia Format: Transition to the Next 100 Years. In: Maschinelle Verarbeitung altdeutscher Texte. Tübingen: Niemeyer, 201-210.

Dignum, F./M Greaves (2000): Issues in Agent Communication. Berlin/Heideberg/New York: Springer.

Fellbaum, Christiane (1998): WordNet: An Electronic Lexical Database. Cambridge, Mass.: MIT Press.

Hausser, Roland (1998): Foundations of Computational Linguistics. Man-Machine Communication in Natural Language. Berlin-Heidelberg_New York: Springer.

Hausser, Roland, ed. (1996): Linguistische Verification. Dokumentation zur Ersten Morpholympics 1994. Tübingen: Niemeyer, 1996.

Hutchins, John (2000): Compendium of Translation Software. Commercial machine translation systems and computer-aided translation support tools. First Edition. Published on behalf of The International Association for Machine Translation by The European Association for Machine Translation. Genève.

Jurafsky, Danie./James H. Martin (2000): Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice-Hall., Inc.

Lenders, Winfried (1974): Semantische und argumentative Textdeskription. Ein Beitrag zur Simulation sprachlicher Kommunikation. Hamburg: Buske.

Sparck Jones, Karen/Martin Kay (1973): Linguistics and Information Science. Academic Press.

Udelhoven, Tanja (2001): RST-basierte Zusammenfassung von Texten. Master-Thesis. University of Bonn.

Wolters, Maria (2001): Towards Entity Status. Phil. Diss. Bonn.