

White Paper on Natural Language Processing

Ralph Weischedel, Chairperson
BBN Systems and Technologies Corporation

Jaime Carbonell
Carnegie-Mellon University

Barbara Grosz
Harvard University

Wendy Lehnert
University of Massachusetts, Amherst

Mitchell Marcus
University of Pennsylvania

Raymond Perrault
SRI International

Robert Wilensky
University of California, Berkeley

1. Scope

1.1. Major Challenges

We take the ultimate goal of natural language processing (NLP) to be the ability to use natural languages as effectively as humans do. Natural language, whether spoken, written, or typed, is the most natural means of communication between humans, and the mode of expression of choice for most of the documents they produce. As computers play a larger role in the preparation, acquisition, transmission, monitoring, storage, analysis, and transformation of information, endowing them with the ability to understand and generate information expressed in natural languages becomes more and more necessary. Some tasks currently performed by humans cannot be automated without endowing computers with natural language processing capabilities, and these provide two major challenges to NLP systems:

1. *Reading and writing text*, applied to tasks such as message routing, abstracting, monitoring, summarizing, and entering information in databases, with applications, in such areas as intelligence, logistics, office automation, and libraries. Computers should be able to assimilate and compose extended communications.
2. *Translation*, of documents or spoken language, with applications, in such areas as in science, diplomacy, multinational commerce, and intelligence. Computers should be able to understand input in more than one language, provide output in more than one language, and translate between languages.

The dominance of natural language as a means of communication in a broad range of interactions among humans suggests that it would be an attractive medium in human-computer interaction as well. The case is particularly strong where the environment precludes the use of keyboard, display, and mouse, so that spoken natural language is almost the only alternative. However, speech recognition alone will not suffice in these settings. Words and phrases must be parsed and interpreted so that their intended meaning (as command, query, or assertion) may be determined and an appropriate response formulated and expressed.

Even where other devices are available, the artificial languages they give the user access to -- menu and icon selection, and programming, command, and database query languages -- are limited. Menus and icons make it easy to present the user with the available options at any time, but they constrain the user to operating on visible objects only. It is also awkward or impossible to operate on sets of objects selected by complex properties (e.g., "Send all the C3 ships with RR1 radar to the nearest port"). Programming and other "linear" artificial languages offer well-defined control structures but are more difficult to learn, and do not take advantage of pointing to objects on the screen. Moreover, they typically require substantial knowledge of underlying representations, and they offer the user little guidance as to what can be done next. In fact, interaction with computers in any artificial language places on the user most of the burden of discovering how to express in the language the commands necessary to achieve the desired objective. It is natural for humans using natural languages to state complex conditions, to integrate these with pointing, and to negotiate how a task could and should be done. Natural language should, however, be seen as a powerful addition to the repertoire of methods for human-machine interaction, and not as a replacement for those methods.

Thus, the third major challenge of natural language processing is

3. *Interactive dialogue*, allowing humans simple, effective access to computer systems, using natural language and other modalities, for problem-solving, decision-making, and control. Application areas include database access, command and control, factory control, office automation, logistics, and computer-assisted instruction. Human-machine interaction should be as natural, facile, and multi-modal as interaction among humans.

While it will be quite some time before systems meet these challenges with the depth and flexibility that humans bring to them, useful shorter term goals of economic value have been and can continue to be met. These include database access systems, multi-modal interfaces to expert systems and simulators, processing of text (e.g., for database update), and semi-automatic machine translation systems.

Virtually all developments necessary to long- and short-term progress toward these goals will be relevant to the transition from speech recognition systems to spoken language systems (SLS), which is the subject of a separate report to this committee. However, in this report we will not have anything further to say specifically about SLSs or their attendant problems in speech signal processing.

1.2. Barriers to Progress

1.2.1. Success and Limitations Thus Far

NLP systems perform three related functions: analysis (or interpretation) of the input, mapping it into an expression in some meaning representation language (MRL); reasoning about the interpretation to determine the content of what should be produced in response to the user, perhaps accessing information in databases, expert systems, etc; and finally, generation of a response, perhaps as a natural-language utterance or text. In translation systems, the "response" is in a language different from the input language.

The most visible results in NLP in the last five years are several commercially available systems for database question-answering. These systems, the result of transferring technology developed in the 1970s and early 1980s, have been successfully used to improve productivity by replacing fourth-generation database query languages. The following case study illustrates their capabilities: with 8 person weeks, one of these systems was ported to a Navy relational database of 666 fields (from 75 relations) with a vocabulary of over 6,000 (root) words.

Queries from a new user of one of these systems are estimated to succeed 60 to 80% of the time; with use of the system users naturally and automatically adjust to the data that is in the database and to the limits of the language understood by the system, giving a success rate of 80 to 95%, depending on the individual.

The success of these systems has depended on the fact that sufficient coverage of the language is possible with relatively simple semantic and discourse models. The semantics are bounded by the semantics of the relations used in databases and the fact that words have a restricted number of meanings in one domain. The discourse model for a query is usually limited to the table output of the previous answer and the noun phrases mentioned in the last few queries.

The limitations of today's practical language processing technology are summarized as follows:

- Domains must be narrow enough that the constraints on the relevant semantic concepts and relations can be expressed using current knowledge representation techniques, primarily in terms of types and sorts. Processing may be viewed abstractly as the application of recursive tree rewriting rules, including filtering out trees not matching a certain pattern.
- Handcrafting is necessary, even in the grammatical components of systems, the component technology that exhibits least dependence on the application domain. Lexicons and axiomatizations of critical facts must be developed for each domain, and these remain time-consuming tasks.
- The user must still adapt to the machine, but, as the products testify, can do so effectively.
- Current systems have limited discourse capabilities which are almost exclusively handcrafted. Thus current systems are limited to viewing interaction, translation, and writing and reading text as processing a sequence of rather isolated sentences. Consequently, the user must adapt to such limited discourse.

1.2.2. Status of Evaluation in NLP

Natural language processing combines basic scientific challenges with diverse technological needs. Evaluating progress in scientific challenges for NLP is a multifaceted issue, perhaps akin to the problem of evaluating progress in the field of programming languages. While certain aspects of progress in that field can be quantified, this is generally not the case; it is hard to quantify how much better one programming language is than another, for example. Nevertheless, there is still reason to believe that scientific issues are becoming better defined and that progress is being made.

On the other hand, evaluation metrics for technological advances based on the science are being developed and applied. In machine translation, measurements have evolved further than in other areas, and they encompass both range of application (what kind of texts in what domain can be translated) and accuracy of the translation (percentage of sentences that remain semantically invariant, and of those, the percentage that are stylistically acceptable). Natural language interfaces can be evaluated in terms of their habitability; that is, how well and how fast can a user get the task accomplished? How often do first phrasings work? Do subsequent human-machine clarification, or focused rephrasing, yield success? These criteria, however, evaluate performance of the technology in a task domain, rather than the underlying science, independent of task.

It behooves the field to continue refining and applying qualitative as well as quantitative measures of progress in the NLP task domains, and to consider the issue of how to measure and evaluate research in the scientific core, if that is to be done any differently from the standard measures of publication and scholarship of most established scientific fields.

1.2.3. Further Scientific Work

Given the limitations of the current technology discussed in Section 1.2.1, fundamental scientific problems in the following three broad areas should be addressed:

Adequate theories of semantics and discourse. These theories should apply to both generation and interpretation of interactive dialogue and text and must support communication across a range of domains of discourse. In semantics, this means at a minimum that we must have a semantics of words (a lexical semantics) that is independent of the domain and of the application, and that the meaning of a word is easily (semi-automatically) related to the concepts of particular domains and applications. Accounts of the combined use of linguistic and non-linguistic (e.g., pointing) information in interactive dialogues should also be developed. The particular styles of certain sublanguages, e.g., Army Operations Orders, should also be accommodated.

Acquisition of information necessary to understanding and creating communication. This includes both linguistic (e.g., words and grammatical forms) and non-linguistic information (e.g., the semantics of icons) used by real users performing real tasks in a given domain of discourse.

A calculus of partial information. In both single- and multiple-sentence texts and dialogues, generation and interpretation requires combining information from a number of sources, such as morphological, syntactic, semantic, pragmatic, and prosodic information. This is particularly true with novel, errorful, and incomplete expressions. Current systems are limited in the kind of information they bring to bear on interpretation, and the processes by which they do so. Both logical and statistical methods may be further investigated.

1.2.4. Non-technical Barriers

Lack of Leverage. Building an NL system requires an extensive effort over several years. Most researchers lack the resources to produce a complete system and lack access to state-of-the-art software for some components (e.g., parsers, large grammars, task-specific lexicons, knowledge representation systems, semantic interpreters). Having such components would let them concentrate their efforts on novel work, demonstrate a complete system, and test individual component theories. Maximal sharing of components requires that a few common tasks be selected by the community and that appropriate backend systems (databases, simulators, expert systems, etc.) be made widely available. Leverage can be further increased by development and support of key NL components. Collection and dissemination of large linguistic data sets will support development of broad-coverage grammars, better lexicons, systematic evaluation procedures, and statistical measures.

Funding. Overall funding for NLP has been strong from 1984 through 1988. However, DARPA funding in the last several years has increasingly emphasized technology transfer and near-term results. Although this emphasis has had some positive as well as negative results, the overall trend is cause for concern. On the positive side, the focus on shorter term performance has forced the community to focus on the development of prototypes addressing specific tasks in specific domains and to think about evaluation methods and resource sharing. On the negative side, it has left little room for developing the theoretical basis of the next generation of systems. Some of this responsibility has been taken on by other sources (notably the Systems Development Foundation and Japanese industry), but support from the former is coming to an end, and there are obvious reasons for not wanting to depend too heavily on the latter. Given these factors, we have serious concern for future levels of basic research funding.

Training of Researchers. Researchers in NLP need a broad exposure to AI, computer science, linguistics, logic, and increasingly to probability and statistics. It is important that the funding of research projects, in and out of universities, allow for student participation.

1.3. Anticipated Developments

During the next decade, we anticipate several scientific breakthroughs which should bring about impact noticeable to the user community.

1.3.1. Scientific Breakthroughs

Within the next 3 to 10 years we foresee the following scientific breakthroughs:

- Architectures that support coordinating syntactic, semantic, and pragmatic constraints, that deal with partial information, and that understand novel, errorful, and vague forms.
- A robust, task-independent, compositional semantics, including more thorough treatment of problems relevant to major application areas, such as time and tense, adverbs and adjectives, conjunctions and ellipsis.
- Automatic acquisition of substantial grammars and lexicons.
- Parallel algorithms for key processes.
- Computational models of discourse structure and speaker intention adequate to support dialogue participation and text generation.

1.3.2. Technology Transfer

Existing laboratory prototypes coupled with the scientific breakthroughs projected above suggest that in the next decade a new generation of systems, having the properties below, will be available:

- Text analysis systems for automatic database update, in restricted domain areas.
- Interactive problem-solving systems combining NL, pointing, and graphical access to several target systems (e.g., databases, simulators, expert systems); exhibiting extended conversations including clarifications, suggestions, and confirmations; and allowing rapid, low-cost portability from one (constrained) application domain to another.
- Language generation systems producing extended texts in limited applications (e.g., summarization of databases or output and explanations of expert systems' decisions).

2. Background

2.1. Current Assessment

Products. In the decade since 1978, at least eight commercial products for natural language access to databases have been released. Two message processing systems are in daily use, one for the U.S. Coast Guard. In the U.S. alone, four companies offer machine translation systems.

Limitations of current systems. The limitations of the current technology, described in Section 1.2.1. of this paper, can be illustrated by considering NL access to databases, the application that has probably received more support than any other in the U.S. in the last ten years. The nature of the task limits the range of inputs the system can expect to see and the semantic distinctions that need to be reflected in the MRL. Reasoning in relational databases is limited to the operations of relational algebra on purely extensional information, so certain concepts, such as tense and modality, need not be reflected in the MRL either. Limitations on the content of a database can

guarantee that certain interpretation ambiguities will not arise, or that they can often be resolved by simple means. In a geographical database, occurrences of the noun "bank" as a financial institution probably never need be considered at all in interpreting "What are the cities on the left bank of the Rhone?". If countries but not mountains have populations, then in "What is the population of Kenya?", Kenya means the country, not the mountain. By assuming the user will adapt to the system and that NL can substitute for an artificial language, the NL interface treats each question in isolation with only a very general, weak notion of the goal in an utterance.

The availability of these kinds of restrictions has allowed NL database query systems to be successful using relatively simple frameworks in which to encode the necessary knowledge sources (grammars, type and sortal information, lists of mentioned entities) applied in relatively simple ways (parsing, recursive tree transformations). This is not to minimize the effort required to build the grammar, semantic model, etc. for a particular application.

The size of the vocabulary *per se* is not a limiting factor, though it does impact the initial cost of bringing up the NLP. Rather what is limiting is the number of word senses per word in the vocabulary, whether the language involves substantial intersentential effects (discourse structure), and whether the underlying semantics is richer than that of relational databases.

Scientific progress. The scientific progress of the last 10 years can be described in terms of traditional linguistic areas and in terms of task areas.

The main development in syntax has been the shift from grammars including procedural constructs to ones expressed purely declaratively. In contrast to context-free grammars, which use atomic symbols only, these so called *unification grammars* use complex terms instead, with term unification instead of equality checking as the main operation. This has allowed for the use of the same grammars with a range of algorithms, both sequential and parallel, for analysis and generation. Grammar development tools have been written in a variety of unification-based frameworks, and widely distributed. Unification grammars are currently being used to apply syntactic (and some semantic) constraints in speech recognition. Within the family of unification grammars lie the "mildly context-sensitive grammars" -- a class that properly contains the context-free grammars, and allows the expression of observed syntactic constraints not expressible in CFGs, but whose recognition problem is polynomial-time computable.

In semantics, the major aspects of the contribution of sentence structure to meaning are understood and implemented. First steps toward automatically extracting aspects of lexical meaning from machine-readable dictionaries have been taken.

Understanding and generating connected sentences introduces questions such as how texts and dialogues are structured; how this structure affects interpretation, particularly of referential expressions; how the beliefs, intentions, and plans of a speaker are conveyed by what is said and how they constrain what is meant, as well as what appropriate responses are. All these questions have been and continue to be investigated. Underlying logics and algorithms for reasoning about knowledge, belief, intention, and action have been proposed, as have initial computational models for discourse structure and methods for planning and plan recognition for discourse.

Although progress continues to be made on query systems, substantial systems have been developed for other applications, all of which were unexplored 10 years ago. These include several text processing systems supported by DARPA's Strategic Computing Initiative (SCI), ranging from systems with very detailed models of one domain, to more general ones adapted to several domains (e.g., Naval Casualty Reports, RAINFORMS, terrorist reports). Multi-media interactive problem solving systems have been developed for the environment of Navy Fleet Command Center decision making and for factory control. Language generation systems have been implemented to

generate multi-sentence explanations of expert system decisions, object or situation descriptions, and instructions from an expert.

2.2. Relationships To Other Areas

Historically, NLP's strongest interaction has been with other areas of artificial intelligence, e.g., with work in knowledge representation and planning. During the 1980s, collaboration with theoretical linguistics and cognitive science has been growing. Lack of widespread availability of high-performance, parallel computers thus far has limited the algorithms considered; however, efforts in parallelism, including work in connectionist neural network modeling, may grow in the next decade.

Until two years ago, interaction with speech scientists had been minimal since the end of the DARPA Speech Understanding Program in 1976. Progress in natural language processing should contribute directly to spoken language systems. This is true not only where understanding seems to be necessary, e.g., voice commands and requests, translation of speech, etc., but also in speech transcription. The error rate of speech recognition systems is directly correlated with the perplexity of the language to be recognized. Statistical language models in speech transcription have given the lowest perplexity thus far, and therefore, the best performance. Language processing techniques, whether supplemental or in place of current statistical models, offer the potential of providing even lower perplexity due to modeling both local and global constraints, as well as supporting speech applications other than transcription.

3. Research Opportunities

To complement our discussion of ultimate goals and the scientific work needed to achieve them, we here outline some nearer term research objectives and address evaluation issues.

3.1. Scientific Objectives

Acquisition of corpora, grammars, and lexicons. The development of useful systems requires observation of the behavior of potential users of interactive systems under realistic circumstances, and the collection of corpora of typical data for text analysis and machine translation systems. Although we believe it is unlikely that full grammars and lexicons can be induced completely automatically in the near future, useful results may be obtained soon from induction and acquisition techniques based on annotated corpora and machine-readable dictionaries. It is also likely that statistical measures useful for biasing algorithms can be extracted from a handcrafted grammar and a corpus. Approaches that appear promising are 1) the learning of grammatical structures where the input has already been annotated by part of speech and/or phrase structure, and 2) the learning of lexical syntax/semantics from examples and/or queries to the user given some pre-coded domain knowledge.

Increasing expressive power of Meaning Representation Languages. Moving beyond database query systems will require increasing the expressive power of the MRL to include at least modal and higher order constructs. Reasoning tools for modal logics and for second-order logics already exist, but they appear intractable for language processing tasks. Approaches that seem promising include encoding modal constructs in first-order logic, hybrid approaches to representation and reasoning, and approaches to resource-limited or shallow reasoning, such as adding weights to formulae and subformulae.

Reasoning about plans. Recent work on plan recognition (the inference of the beliefs and intentions of agents in context) has provided formal definitions of the problem and some new algorithms. These have not yet

been used as part of a discourse component to help resolve reference, quantification, and modification ambiguities or to formulate an appropriate response. The interaction between plans, discourse structure, and focus of attention also needs to be investigated. Promising approaches include incorporation of beliefs of the discourse participants, integrating existing models into discourse processing under simplifying conditions, and exploring prosodic and linguistic cues to dialogue.

Combination of partial information. The standard control structure by which various sources of information are combined in language interpretation seems to limit what NL systems can do. Several proposals for more flexible control structures have been made recently, each covering a subset of the knowledge sources available. More comprehensive schemes need to be developed. Two promising approaches are generalization of unification to NL architectures and use of global, weighted control strategies, as in evidential reasoning.

Improving robustness. Published studies suggest that as much as 25 to 30% of typed input contains errors, is incomplete, uses novel language, or otherwise involves challenging phenomena that are not well handled theoretically. Some experts believe the frequency of occurrence for these classes is even higher in spoken language than in written language. The text of some messages, such as Navy RAINFORM and CASREP messages and bank telexes, is highly telegraphic. It should be possible to develop a domain-independent theory that allows at least partial understanding of some of these novel and errorful uses, and to test it in narrowly defined domains. Promising approaches are to employ unification strategies, plan recognition, and weighted control strategies to determine the most likely interpretation and the most appropriate response/action.

Exploring the relationship between linguistic and conceptual knowledge. Use and interpretation of metaphor illustrates a prevalent relationship between linguistic and conceptual knowledge. For example, there is obvious systematicity in the use of expressions like "kill the engine" and "my engine died" which extends to other domains ("kill that process" and "my process died"). An understanding of such issues has been shown in the lab to be effective for language learning, where such regularities have been effectively exploited to learn extended word meanings. A domain-independent theory of the relationship could be developed and tested.

Relating interpretation and action. The problem of how to relate interpretations expressed in an MRL and calls to application systems (databases, summarizing algorithms, etc.) has not been fully resolved, or in fact precisely stated. Resolving this relationship is crucial to the systematic separation of the natural language part of the system from the application part. Any approach should deal with applications beyond databases (beyond the semantics of tables) and should avoid the challenges of automatic programming.

Finding the relationship between prosody, syntactic ambiguity, and discourse structure. Syntactic and discourse boundaries are one of the main sources of interpretation ambiguity. Recently discovered evidence shows that prosodic information is a good indicator of these boundaries. Automatic extraction of prosodic information would revolutionize the interpretation of spoken language. Further, generation systems could add prosodic information to signal syntactic structure and discourse structure.

Facilitating leverage through shared resources. To address the problem of lack of leverage, several projects could be funded to support efforts throughout a significant portion of the community. We believe that such projects will make both the conduct and the evaluation of natural language work substantially faster and easier; that they can significantly reduce duplication of effort; that they can help to facilitate the individual researcher's efforts on new work rather than on infrastructure; and that they can materially increase the compatibility of research and development activities at different sites. Examples of infrastructure include:

1. Collection and labeling of several corpora of various genres (dialogues, essays, narratives, etc.). Some experts believe a corpus of 100,000,000 words is required. The labeling should include part of speech,

- syntactic structure information, co-reference, and any semantic/pragmatic information that can be reliably added.
2. Distribution and maintenance of two or more of the most extensive grammars and parsers of English.
 3. Collection of a substantial lexicon with feature information that is uncontroversial. (Since the largest lexicons thus far have been about 10,000 words, the size should be at least 20,000.)
 4. Maintenance of two or more knowledge representation and reasoning systems.
 5. Distribution and maintenance of two or three natural language interfaces.
 6. Distribution of one or more "backend" systems to serve as the target of an interface.

3.2. Measures of Progress

The means of measuring progress is still an active area of discussion among NL scientists, as evidenced by the Workshop on Natural Language Evaluation held outside Philadelphia in December 1988. Measures of correctness can be relatively simply stated for database query systems without dialogue capabilities (e.g., without sequence-related queries or clarifications), or for text analysis systems for database entry. They are much more difficult to state when stylistic matters need to be considered (as in MT systems) or when system responses affect subsequent user utterances. They probably can't be usefully stated in a domain- or task-independent way. Measures of task difficulty, or of ambiguity of the language model, analogous to speech recognition's perplexity, are much more difficult to state.

Measurement of NL systems requires three distinct types of comparisons:

1. Longitudinal: It is critical to be able to measure the performance of a system over time, so that progress can be tracked.
2. Cross-System: It should be possible to compare the overall performance of two systems in explicit terms. This focus on whole-system performance will help localize the strengths and weaknesses of complete systems and will identify topics for research and development efforts.
3. Component: It should be possible to evaluate and compare parts of systems and evaluate coverage of unknown phenomena. This focus on components will help point out areas of relative strength in different systems and will provide priorities and goals for specific research.

4. Impact

4.1. Potential Impact

One way to assess potential impact is via a holistic, subjective view. The following is a summary of a market survey that uses this approach:¹

Considering the question of feasibility (on NLP systems) first, the answer must be "yes" ...

The second - easier - issue, is whether people will really want computer NLP. In the style of some systems of logic, it can be resolved by testing for the negative; under what circumstances would people not want computers to handle natural language, given that they can satisfy price and performance requirements? The only obvious answers are: where a routine action (such as pressing a button or keying in a standard command) is quicker or more convenient; where communication in a strictly formalized language is more reliable, precise, or subtle; and where there are overriding requirements for human involvement, ranging from legal obligations to job preservation. These

¹Tim Johnson, *Natural language computing: the commercial applications*, Ovum LTD, London, 1985, pp. 45-46

considerations will rule out quite a number of possible applications, but the broad conclusion must be that if the technology is available, then it will be widely used.

We illustrate the conclusion above in the following three subsections. A second way to view impact is to quantify the potential marketplace. We report on this approach in Section 4.1.4.

4.1.1. Human-machine Interaction Systems

It is difficult to overestimate or overstate the technical, scientific, and socio-economic value of an effective and efficient means of human-machine interaction. Even so, it is easy to lose sight of the importance of the interface in the effort to develop some underlying functionality. Indeed, it is worth pointing out that the ideal interface is "invisible," in the sense that users find it so natural and easy to use that they are never aware of the interface itself. Such interfaces will make computers usable by everyone, without special training. Components of this work could be incorporated into virtually every human-machine system developed or supported by the government. Just as independent menu, graphic, text, and speech I/O capabilities have evolved from DARPA-supported work, so will the integrated, multimodal problem-solving environments made possible by work on interfaces become commonplace.

4.1.2. Reading And Writing Text

Virtually all workplaces are inundated by documents, forms, messages, memos, and reference archives. With far more computing power and memory at the fingertips of office staff, with the rapid growth in networks, and with the demand for timely information, it is not wild speculation to foresee a crisis of information gridlock in whole organizations. Those who can digest the necessary information first will have the advantage in the economic, political, and military battles of the future. The problem is not merely one of ingesting new information, though; productivity in retrieving, editing, maintaining, and creating information is also critical.

Consider the potential in the area of intelligence analysis. Impressive improvements in the means of collecting data for intelligence analysis have far outstripped advances in technology to help the intelligence analyst use the data collected. It is now possible to collect more free text reports than can possibly be digested. The reports come in many languages, not just English. The availability of reports is going to grow further in the 1990s, but there is no hope for commensurate personnel increases to deal with the availability of data. Natural language processing seems the only hope for aiding in the selection, prioritization, filtering, and analysis of data.

The kinds of aids or utilities that would provide help are broad in scope:

- *Language identification.* Given a segment of free text, identifying the languages it is written in.
- *Prioritization.* Given a message (in free text), assigning a priority to it, based on message content.
- *Routing.* Determining which offices should receive a copy of the text based on its content.
- *Gisting.* Automatically adding records to a database, given the content of free text.
- *Fusion.* Recognizing that a new piece of text correlates with previously known information. Identification of what is new in the message, what corroborates known data, and what conflicts with known data.
- *Report generation.* Automatic preparation of text and tables describing a message, set of messages, or situation.
- *Alerts.* Given some pre-defined criteria about the content of a knowledge base/database, sending a message notifying that the criteria have now been met.

4.1.3. Machine Translation (MT)

The need for extensive translation capabilities, whether human or machine, is becoming increasingly important to the U.S., because of both the increasing importance of world markets to U.S. business and the increasing role of joint military operations. Aids to translation therefore can help the U.S. achieve success in an evolving world. The potential impact of MT is indicated below by a quote:²

Today, more than 20 years after computerized language translation was laughed out of the funding process in the United States, several Japanese companies and industry/government collaborations are beginning to turn the once-derided technology into a gold mine of new applications and opportunities. The top U.S. company estimates that the annual market for international translation is at least \$10 billion, and as machine translation systems improve they will command an increasing share of a market growing at a rate of 10 to 15 percent a year.

The key to why there is a market for MT today is the fact that one need not have fully automatic high-quality translation to have a valuable product; tools that increase productivity are sufficient. If economies are gained by editing a translation drafted using MT, that is sufficient to warrant use of MT systems.

In terms of R&D funding, MT is the application of NLP that is attracting the most funding from government and industry in Japan and Europe. In Europe, Eurotra alone is spending \$20 million in MT, with another expected \$20 million in matching funds. The total Japanese investment is even larger (most of it industrial). ATR (in Osaka, funded by the ministry of postal and telecommunication) is doing a 15-year project on simultaneous "interpreting telephony" combining MT, dialog analysis, and speech.

4.1.4. Forecasting the Market

Though forecasting the market for a technology that is emerging from the laboratory is not a reliable process, we cite what we feel is the most useful market survey. Figure 4-1 shows estimated sales in 1987 of NLP systems in the U.S.; Figure 4-2 shows estimated sales of NLP products in 1985 and 1987 and forecasted sales for 1989, 1991, 1993, and 1995. In the diagrams, "content scanning" corresponds to the ultimate goal of text reading, and "talkwriter" corresponds to the goal of automatic speech transcription.

4.2. Transfer to the Real World

Transfer to the real world is already occurring in three application areas: database retrieval, message processing in highly constrained domains, and aids to document translation. As can be seen from the forecasts in the previous section, prospects for continued transfer of technology seem bright. Nevertheless, government support for particular aspects will be necessary to develop the technology and its commercialization in ways that might otherwise be long delayed. For instance, multi-modal interfaces that include both text and speech processing are clearly mandated in military applications, where high-quality audio-visual hardware is expected; on the other hand, market forces can argue for supporting the lowest common denominator in hardware, such as, a terminal, because it is the most dominant screen technology currently commonplace. Similarly, commercial sources can be expected to fund evolutionary improvements in current technology and short-term risks likely to have high payoff. However, the next one or two generations of science and technology cannot be expected to emerge without substantial government funding.

²R. C. Wood, "The Language Advantage: Japan's Machine Translators Rule the Market", *High Technology Business*, November 1987, p. 17.

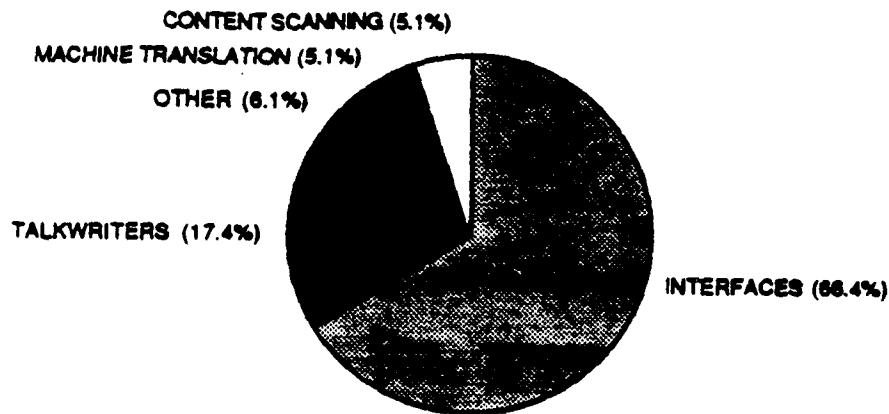


Figure 4-1: NLP Applications--Market Share, USA, 1987

<u>Area</u>	<u>1985</u>	<u>1987</u>	<u>1989</u>	<u>1991</u>	<u>1993</u>	<u>1995</u>
Interfaces	12.1	21.7	36.3	64.4	137.4	254.4
Machine Translation	2.3	1.7	2.9	8.2	9.6	14.4
Content Scanning	0.7	1.7	5.1	11.0	20.4	39.1
Talkwriter	0.0	5.7	22.0	68.8	177	306
Other	0.5	2.0	2.0	9.4	25.4	90.0

Figure 4-2: Forecasts for Natural Language Products by Application in Millions of Dollars³

We believe a secondary effect of the Strategic Computing Program has been greater industrial R&D effort in natural language technology. Government investment in technology transfer can further prime the pump of industrial investment.

Judging when a laboratory system is ready for transfer to the real world is difficult. One approach to evaluating whether a system is ready is to measure the effort required to achieve some specified level of performance in a new application domain. Such a measure indicates not only the cost of applying the system but also its degree of maturity. The reason such measures are critical is that a domain-independent lexical semantics and domain-independent discourse processing are areas in which further scientific research is needed.

³Tim Johnson, "Commercial markets for natural language processing", talk presented at the Second Conference on Applied Natural Language Processing, Association for Computational Linguistics, February, 1988.

5. Conclusions and Recommendations

The impact of a breakthrough in computer use of natural languages will have as profound an effect on society as would breakthroughs in superconductors, inexpensive fusion, or genetic engineering. The impact of NLP by machine will be even greater than the impact of microprocessor technology in the last 20 years. The rationale is simple: natural language is fundamental to almost all business, military, and social activities; therefore, the applicability of NLP is almost limitless.

NL analysis and generation could revolutionize our individual, institutional, and national ability to enter, access, summarize, and translate textual information. It can make interaction with machines as easy as interaction between individuals.

The computer's linguistic proficiency may never be as great as a human's. However, the existence and use of current NL products and the market projections cited suggest that investment in this technology should lead to useful spinoffs in the near term and mid-term.

The technology stands at a turning point. New approaches (see Section 3.1) offer opportunities for substantial progress in the next five years, and breakthroughs within 3 to 10 years.

Given these conclusions, we have three recommendations:

1. Support both component research and system integration designed to achieve successes in the near term, i.e., the scientific breakthroughs and technology transfer projections made in Sections 1.3.1 and 1.3.2. Some of these projects should produce demonstrable results in applications including (but not restricted to) machine translation, interactive dialogue systems for problem solving and consulting, and text input/output.
2. Invest in approaches to the areas labeled further scientific work in Section 1.2.3, particularly in high-payoff approaches. The goal is the creation and fostering of seminal ideas that could lead to long-term breakthroughs.
3. Support infrastructure to leverage research, such as large annotated corpora, very large grammars, theory-neutral lexicons containing tens of thousands of words, common-sense knowledge bases, modular NLP systems, and application backends.
4. Increase overall funding for NL research, since two new challenges face the U.S. First, the need for NL processing to support intelligence analysis is already clear, will only grow in the next decade, and has not been addressed by previous DARPA programs.⁴ Second, Japanese successes in machine translation of text and the Japanese emphasis on simultaneous translation (of speech) suggest the desirability of a program that supports approaches to machine translation that offer promise of scientific breakthroughs and progress on the long-term objectives identified in Section 1: reading and writing text, translation, and interactive dialogue.

⁴Previous DARPA programs have focused on English, whereas the needs in the intelligence community embrace several critical languages.