Natural language processing: a historical review

Karen Sparck Jones
Computer Laboratory, University of Cambridge
William Gates Building, JJ Thomson Avenue, Cambridge CB3 0FD, UK
sparckjones@cl.cam.ac.uk, http://www.cl.cam.ac.uk/~ksj

October 2001

Abstract

This paper reviews natural language processing (NLP) from the late nineteen forties to the present, seeking to identify its successive trends as these reflect concerns with different problems or the pursuit of different approaches to solving these problems and building systems as wholes. The review distinguishes four phases in the history of NLP, characterised respectively by an emphasis on machine translation, by the influence of artificial intelligence, by the adoption of a logico-grammatical style, and the use of massive language data. The account considers the significant and salient work in each phase, and concludes with an assessment of where we stand after more than fifty years of effort in the field. [1]

1 Introduction

Work in the NLP field has concentrated first on one problem, then on another, sometimes because solving problem X depends on solving problem Y but sometimes just because problem Y seems more tractable than problem X, or because there is market interest in a solution to Y. There has been very substantial progress, both in understanding how to do NLP and in actually doing it, since work in the field took off in the 1950s. In the last fifteen years in particular, advances in computing technology have made it possible to implement ideas that could only be adumbrated before, to consolidate research, and to carry speech and language processing into the ordinary world. Sometimes the scientific advance in NLP, or the computational linguistics underlying it, is less than the onward rush of information technology is taken to imply; but the theoretical and empirical development needed to establish a new field is evident in the fifty-year period reviewed here. Sometimes innovation is only old ideas reappearing in new guises, like lexicalist approaches to NLP, or shallow parsing. But the new costumes are better made, of better materials, as well as more becoming: so the research is not so much going round in circles as ascending a spiral.

In this brief review, I divide the history of NLP into four phases, with distinctive concerns and styles. I define the first phase of work in NLP as lasting from the late 1940s to the

[1] This paper uses material from my article 'Natural language processing: an overview', in the first and second editions of *International encyclopedia of linguistics*, New York: Oxford University Press, 1992, (Ed. W. Bright), and in press (Ed. W. Frawley), and from 'Natural language processing: a historical overview' in *Current issues in computational linguistics: in honour of Don Walker*, Amsterdam: Kluwer, 1994, (Ed. A. Zampolli, N. Calzolari and M. Palmer).

late 1960s, the second from the late 60s to the late 70s and the third to the late 80s, with the fourth phase to the end of the century. The first phase was driven by MT, the second flavoured by AI, the third grammatico-logical, while the fourth phase has focused on lexical and corpus data. This last phase illustrates the cyclic character of work in the field, since it has returned to a major concern of the first period, namely building the powerful and comprehensive dictionaries that serious NLP applications like MT need, and has revived early proposals for extracting linguistic information from large natural corpora. At the same time, it also illustrates the extent to which real progress has been made in the field in such key areas as computational grammar and parsing.

2 Phase 1: late 1940s to late 1960s

The work of the first phase was focused on machine translation (MT). Following a few early birds, including Booth and Richens' investigations and Weaver's influential memorandum on translation of 1949 (Locke and Booth, 1955), research on NLP began in earnest in the 1950s. Automatic translation from Russian to English, in a very rudimentary form and limited experiment, was exhibited in the IBM-Georgetown Demonstration of 1954. The journal *MT (Mechanical Translation)*, the ancestor of *Computational Linguistics*, also began publication in 1954. The first international conference on MT was held in 1952, the second in 1956 (the year of the first artificial intelligence conference); at the important Washington International Conference on Scientific Information of 1958 language processing was linked with information retrieval, for example in the use of a thesaurus, Minsky drew attention to artificial intelligence, and Luhn provided auto-abstracts (actually extracts) for one session's papers. The Teddington International Conference on Machine Translation of Languages and Applied Language Analysis in 1961 was perhaps the high point of this first phase: it reported work done in many countries on many aspects of NLP including morphology, syntax and semantics, in interpretation and generation, and ranging from formal theory to hardware.

This first phase was a period of enthusiasm and optimism. It is notable not only because those engaged attacked a very difficult NLP task, and so encountered the problems of syntactic and semantic processing, and of linguistic variety, in all their force; they were seeking to use a new tool, computers, for non-numerical, data-processing purposes when data-processing itself was not well established. It is essential to remember how primitive the available computing resources were. This was the era of punched cards and batch processing. There were no suitable higher-level languages and programming was virtually all in assembler. Access to machines was often restricted; they had very limited storage, and were extremely slow. Plath (1967) reports processing speeds like 7 minutes for analysing long sentences, even with the most advanced algorithms and on the best machines then available. Vast amounts of programming effort were devoted to bit-packing to save space and time. It is remarkable how much was done with such poor resources, for example in grammar and lexicon building: some of the grammars and dictionaries of the early 1960s were very large even by current standards.

Research in this period was thoroughly international, with considerable activity in the USSR as well as in the USA and Europe, and some in Japan. US grant funding increased after Sputnik 1, but the work had begun before. Russian and English were the dominant languages, but others, including Chinese, were involved (Booth, 1967; Hutchins, 1986, 2000).

Though the period ended under the cloud of the 1966 ALPAC Report, most of those engaged were neither crooks nor bozos. Many came to NLP research with a background

and established status in linguistic and language study, and were motivated by the belief that something practically useful could be achieved, even though the strategies adopted were crude and the results not of high quality. The first major question was whether even to obtain only limited results, principled methods based on generalisation were required, or whether ad hoc particularisation would suffice. The second issue was the relative emphasis to be placed, in either case, on syntax and on semantics. The third problem was the actual value of the results, especially when balanced against pre- or post- editing requirements.

The main line of work during this period can be summarised as starting with translation as lookup, in dictionary-based word-for-word processing. The need to resolve syntactic and semantic ambiguity, and the former in particular because it is not open to fudging through the use of broad output equivalents, led to ambiguity resolution strategies based on local context, so dictionary entries became in effect individual procedures. Semantic resolution involved both specific word, and semantic category, collocation. But long-distance dependencies, the lack of a transparent word order in languages like German, and also the need for a whole-sentence structure characterisation to obtain properly ordered output, as well as a perceived value in generalisation, led to the development of autonomous sentence grammars and parsers.

Most of the NLP research done in this period was focused on syntax, partly because syntactic processing was manifestly necessary, and partly through implicit or explicit endorsement of the idea of syntax-driven processing. The really new experience in this work, and its contribution to linguistics in general, came from recognising the implications of computing represented by the need not only for an explicit, precise, and complete characterisation of language, but for a well-founded or formal characterisation and, even more importantly, of the need for algorithms to apply this description. Plath's account (1967) of NLP research at Harvard shows this development of computational grammar with its lexicon and parsing strategy very clearly. But as Plath also makes clear, those concentrating on syntax did not suppose that this was all there was to it: the semantic problems and needs of NLP were only too obvious to those aiming, as many MT workers were, at the translation of unrestricted real texts like scientific papers. The strategy was rather to tackle syntax first, if only because semantic ambiguity resolution might be finessed by using words with broad meanings as output because these could be given the necessary more specific interpretations in context.

There were however some workers who concentrated on semantics because they saw it as the really challenging problem, or assumed semantically-driven processing. Thus Masterman's and Ceccato's groups, for example, exploited semantic pattern matching using semantic categories and semantic case frames, and indeed in Ceccato's work (1967) the use of world knowledge to extend linguistic semantics, along with semantic networks as a device for knowledge representation.

MT research was almost killed by the 1966 ALPAC Report, which concluded that MT was nowhere near achievement and led to funding cuts especially in the most active country, the USA, even though it recommended support for computational linguistics. But it is important to recognise what these first NLP workers did achieve. They recognised, and attempted to meet, the requirements of computational language processing, particularly in relation to syntactic analysis, and indeed successfully parsed and characterised sentences. They investigated many aspects of language, like polysemy, and of processing, including generation. They addressed the issues of overall system architectures and processing strategies, for example in direct, interlingual or transfer translation. They began to develop formalisms and tools, and some influential ideas first appeared, like the use of logic for representation (cf Yngve, 1967). Some groups were also established, developing resources like grammars and gaining

experience, as at the Rand Corporation. There was indeed enough knowhow by now for some textbooks, like Hays (1967).

There was little work, on the other hand, on some important problems that have since attracted attention, like anaphor resolution, since though text was being translated it was treated as a sequence of independent sentences, or on the function of language, since the work was mainly on single-source discourse. There was little attempt to incorporate world knowledge, and to relate this non-linguistic to linguistic knowledge, though some world knowledge was smuggled in under the heading of semantics. The belief, or challenge, was that one could get far enough with essentially linguistic, and therefore shallow, processing not involving reasoning on world models. The research of this period did not produce any systems of scope or quality, though by the end of the 1960s there were MT production systems providing output of use to their customers (Hutchins, 1986). There was more merit in the work of the period, and more continuity, through individuals, with later effort, than subsequent myths allow, though the early literature was inaccessible and little used. But perhaps the best comment is Bledsoe's at the International Joint Conference on Artificial Intelligence of 1985 (Bledsoe, 1986) on the value, for artificial intelligence as a whole, of the early MT workers' head-on attempt to do something really hard.

Work on the use of computers for literary and linguistic study also began in this period, but it has never been closely linked with that in NLP, though some common concerns have become more prominent recently.

3 Phase 2: late 1960s to late 1970s

The second phase of NLP work was artificial intelligence (AI)-flavoured, with much more emphasis on world knowledge and on its role in the construction and manipulation of meaning representations. Pioneering work influenced by AI on the problems of addressing and contructing data or knowledge bases began as early as 1961, with the BASEBALL question-answering system (Green et al., 1961). The actual input to these systems was restricted and the language processing involved very simple compared with contemporary MT analysis, but the systems described in Minsky (1968), and Raphael's SIR in particular, recognised and provided for the need for inference on the knowledge base in interpreting and responding to language input.

Woods et al.'s LUNAR (Woods, 1978) and Winograd's SHRDLU (Winograd, 1973) were the natural successors of these systems, but they were widely seen at the time as representing a step up in sophistication, in terms both of their linguistic and their task-processing capabilities. Though differing in many ways they shared a procedural style and were perceived as having an overall coherence as systems and a genuinely computational character. The dominant linguistic theory of the late 1960s, transformational grammar, was seen both as fundamentally unsuited to computation and particularly analysis, even though TG was formally oriented and there was at least one serious transformational parser, and as offering nothing on semantics, which had to be tackled for any actual NLP system. The computational confidence illustrated by Woods' and Winograd's work, and the range of experiment it promoted, while drawing on previous work, is well shown by the varied research reported in Rustin (1973).

The view that current linguistics had nothing to contribute, and the feeling that AI was liberating, were also apparent in Schank's work (1980), which explicitly emphasised

semantics in the form of general-purpose semantics with case structures for representation and semantically-driven processing. The community's concern, illustrated by Winograd and Schank alike, with meaning respresentation and the use of world knowledge then became an argument, reflecting a widespread feeling in AI stimulated by Minsky's promulgation of frames, for the use of a larger scale organisation of knowledge than that represented in NLP by verb case frames or propositional units: this large-scale organisation would characterise the different relationships between the elements of a whole universe of discourse, and would support the inferences, including default inferences, needed especially in interpreting longer discourse and dialogue. NLP would deliver deep representations integrating and filling out individual inputs to form a whole constituting an instantiation of a generic world model. Schank's arguments for the Yale group's use of more event-oriented scripts developed this line in the context of earlier work by linking individual propositional case frames with the larger structures via their semantic primitives (cf Cullingford, 1981). Semantic networks (Bobrow and Collins, 1973; Findler, 1979) were similarly proposed as a third variant on this theme, offering a range of options from associative lexical networks only weakly and implicitly embodying world knowledge to alternative notations for frames. These types of knowledge representation linked NLP with mainstream AI, and their descriptive and functional status, for example in relation to logic, was and has remained a matter for debate.

Semantic primitives seen, as in Schank's Conceptual Dependency, as having a representational and not just a selective role also appeared to fit naturally with the need to capture underlying conceptual relations and identities in discourse processing, particularly for types of material or task where fine distinctions do not figure. Their status too was a matter for controversy, but they have continued in use, supplemented by or sometimes in the form of domain-specific categories, in application systems. They have also had a significant role, in the more conventional form of selectional restrictions, even when semantic driving has been abandoned.

The general confidence of those working in the field, and the widespread belief that progress could be and was being made, was apparent on the one hand in the ARPA Speech Understanding Research (SUR) project (Lea, 1980) and on the other in some major system development projects building database front ends. Several of the SUR projects were ambitious attempts to build genuinely integrated systems combining top-down with bottom-up processing, though unfortunately the best performing system against the target measurements was the least theoretically interesting.

The front end projects (see e.g. Hendrix et al., 1978) were intended to go significantly beyond LUNAR in interfacing to large autonomous (and therefore not controlled) databases, and in being more robust under the pressures of 'ill-formed' input; and the confidence on which they were based drove other work including that on the first significant commercial front end, INTELLECT (Harris, 1984). But these projects unfortunately also showed that even an apparently straightforward, and perhaps the simplest because naturally constrained, NLP task was far more difficult than it seemed to be. NLP workers have been struggling ever since on the one hand with the problems of constructing general-purpose transportable front ends and of providing for the acquisition of application-specific knowledge, and on the other of handling the user's real needs in dialogue. The former led to the development of modular architectures, general-purpose formalisms, and toolkits, typically for supplying a specialised lexicon, semantics, and domain and database model on top of standard syntax, following the sublanguage approach which had been pioneered for text processing by Sager's NYU group (in Kittredge and Lehrberger, 1982), but sometimes supplying a specialised syntax as well.

The latter stimulated research on the identification of the user's beliefs, goals and plans which is also and more fully needed for dynamic and extended interaction with expert systems for consultation and command, where the system's responses should be cooperative.

The need to identify the language user's goals and plans was early recognised by the Yale group, and has become a major trend in NLP research since, along with a more careful treatment of speech acts. Work on interactive dialogue in particular, from the second half of the 70s, has emphasised the communicative function of language, and the indirect function and underlying meaning, as well as direct function and surface meaning, of linguistic expressions. At the same time work on discourse understanding in the 70s, whether on single-source texts like stories or reports, or on dialogue, stimulated research on anaphor resolution and on the construction, maintenance and use of discourse models not relying only on prior scenarios like scripts; and some useful progress was made with the development of notions of discourse or focus spaces and of resolution algorithms tied to these (Joshi et al., 1981; Brady and Berwick, 1983; Grosz et al., 1986, Sections III and IV).

4 Phase 3: late 1970s to late 1980s

It was nevertheless apparent by the early 1980s that it was much harder to build well-founded, i.e. predictable and extensible, NLP systems even for heavily restricted applications than had been supposed, and that systems for more challenging applications in terms of processing tasks or discourse domains could not generally be built in an ad hoc and aggregative way, though claims were made for this as a possible strategy for utilitarian MT, given enough investment of effort.

If the second phase of NLP work was AI-flavoured and semantics-oriented, in a broad sense of "semantic", the third phase can be described, in reference to its dominant style, as a grammatico-logical phase. This trend, as a response to the failures of practical system building, was stimulated by the development of grammatical theory among linguists during the 70s, and by the move towards the use of logic for knowledge representation and reasoning in AI. Following augmented transition networks as computational grammars in a theoretical as well as practical sense, linguists developed a whole range of grammar types, for example functional, categorial, and generalised phrase structure which, because they are oriented towards computability as an abstract principle, are also relevant to actual parsing, particularly since they also tend to have a context-free base supporting efficient parsing algorithms. The emphasis was also on a declarative approach and on unification as the fundamental process, which fitted naturally with a general trend in computing in this period associated with, for example, the growth of logic programming. The processing paradigm, for analysis in particular, was therefore syntax-driven compositional interpretation into logical forms.

Computational grammar theory became a very active area of research linked with logics for meaning and knowledge representation able to deal with the user's beliefs and intentions, and with discourse features and functions like emphasis and theme. Capturing the refinements of linguistic expressions indicating time and mood, say, and of large-scale discourse and dialogue structure remain challenges for this paradigm; but the grammatico-logical approach led by the end of the decade to powerful, general-purpose sentence processors like SRI's Core Language Engine (Alshawi, 1992), and Discourse Representation Theory (Kamp and Reyle, 1993) offered a means of tackling more extended discourse within the grammatico-logical framework.

The grammatico-logical approach also encouraged a more general, if informal, use of predicate calculus-style meaning representations and led, via work on system building, to a shift in the meaning of "semantic" and "pragmatic" and a redistribution of effort within a system as a whole. The linguistic process of semantic interpretation was restricted, and the full meaning of expressions could only be derived by reference to the (application specific) pragmatic context, which subsumed both prior discourse context and the application's domain or world model.

Altogether the period was one of growing confidence and consolidation, and also an expanding community. Practical resources, e.g. grammars, and tools, e.g. parsers, became available (e.g. the Alvey Natural Language Tools (Briscoe et al., 1987)); there were more operational and commercial systems, e.g. for database query; there was a conspicuous revival of work on MT, especially in Europe and Japan, often with a heavy application-specific emphasis; and new tasks were addressed e.g. message understanding (information extraction). These developments were reflected by the ACL's Applied NLP Conferences from 1983, and were significantly promoted by the beginning of the US technology development and evaluation programmes in speech and language processing. The (D)ARPA speech recognition and message understanding (information extraction) conferences were important not only for the tasks they addressed but for their emphasis on rigorous evaluation, initiating a trend that became a major feature of the 1990s (Young and Chase, 1998; Sundheim and Chinchor, 1993).

There was also, following the early work of the 1970s, a practically-motivated surge of research on discourse, especially dialogue, and on generation, particularly multi-sentence text generation. These were connected because cooperative response, e.g. in advice giving systems, depends on modelling the participant's beliefs, goals and plans, and can naturally lead to paragraph-length outputs, for instance in providing explanations. Work on user modelling (Kobsa and Wahlster, 1989) was one strand in research on language processing for such active communication and on discourse structure serving this (Cohen et al., 1990). At the same time, as McKeown (1985) showed, rhetorical schemas could be useful recipes for producing both linguistically coherent and communicatively effective text.

Outside the main line, research on connectionist approaches to NLP signalled important topics for the future, notably word sense disambiguation (Small et al., 1988) and, via probabilistic networks, statistically-coloured NLP. The final trend of the 1980s, work on the lexicon, also pointed in this direction. This was stimulated by the important role the lexicon plays in the grammatico-logical approach, by MT requirements, and by practical needs for transportable or readily customisable systems. Serious attempts were made to exploit commercial dictionaries in machine-readable form, which in turn led to work using test corpora to validate or enhance existing lexical data, and also to projects to define formalisms, e.g. exploiting feature structures, that can both capture rich lexical information and permit its direct use, including inferential use, within a formal processing context.

5 Phase 4: The 1990s

The lexicalist approach to grammar that appeared in the 1980s has become increasingly influential. The lexicon has taken over much of what was formerly assigned to the syntactic component, leaving the latter with few, general rules. Parsing with head phrase structure grammars, for instance, is a compositional operation on large blocks of constituent feature data, with corresponding semantic rules delivering logical forms as before. Lexicalised tree

adjoining grammars illustrate the same trend. However there has been a very competitive revival, stimulated by practical tasks like information extraction from news material, of simple finite state methods; and formal linguistic modelling has come under more general attack from statistical approaches to data characterisation and processing.

Statistical language processing has indeed been the most striking feature of the last decade (Manning and Schuetze, 1999). This has involved not only data analysis, e.g. for semantic classification, an old idea rejuvenated by the arrival of vast quantities of machine-readable data and the machine power to handle it (Sparck Jones, 1992). It has more importantly involved the direct application of statistical methods to NLP itself, e.g. for probabilistic parsing. The two are combined in machine learning from corpus data, say to derive both syntactic rules and their probabilities. This development was partly a response to the challenges presented by the rich, varied and unpredictable material, like news text, to which e.g. translation systems were to be applied and for which absolutist approaches are manifestly unsuited. It was also a response to the successes of Hidden Markov Modelling in speech recognition.

The third feature of the 1990s has been significant progress with practical tasks, especially those for which shallow processing methods, e.g. using finite state parsing and surface patterns, can deliver useful output. The vast quantities of text flooding the World Wide Web have in particular stimulated work on tasks for managing this flood, notably by information extraction and automatic summarising (Mani and Maybury, 1999), and on sufficient-to-the-day strategies for these tasks, for instance text extraction using word frequency and discourse structure cues. These tasks have also, for the first time, seriously involved spoken language. Speech recognition (transcription) technology has advanced rapidly and natural, albeit very simple, speech dialogue systems are now commercial realities, and more ambitious ones involving translation like Verbmobil (Wahlster, 2000) are beginning to appear. The shallow processing approach has been extended to dialogue structure through the use of conversational games rather than intensional user modelling, and relatively straightforward rules have been found surprisingly effective for purposes like focus determination and anaphor resolution. These developments (cf. Cole et al., 1998) are reflected in the use of "language engineering" as distinct not only from "computational linguistics", but also "natural language processing".

This emphasis on language technology has been naturally associated with the fourth major trend of the decade, namely evaluation (Sparck Jones and Galliers, 1996). The US Government programmes, extended over an increasing range of tasks, have played a significant part, through carefully designed and challenging evaluations, in stimulating work on generic task systems and general-purpose components. They have placed particular emphasis on robustness and portability, i.e. on systems that either do not require heavy domain knowledge or can be easily customised. Similar European initiatives have focused largely on multilingual tasks and issues. This concern with evaluation has been associated with a conspicuous growth of interest in the design and provision of linguistic resources, for example the British National Corpus (see website reference) and WordNet (Fellbaum, 1999), and test tools like the Penn Treebank, which has annotations allowing parser performance assessment (see website). Public domain processing tools, e.g. taggers, are also now available, allowing rapid prototype system assembly using modular architectures.

6 The present

Natural language processing is currently flourishing. Information technology developments are encouraging work on a broad range of tasks designed to find, digest, package, and represent speech or text-derived information. MT (if only of a kind), for instance, is now a routine offering by search engines, and more useful, because better, NLP systems will appear. At the same time, foundational work, seeking to combine the formally sound with the computationally real, continues. One important aspect of this, that Sparck Jones et al. (2000) marks as a trend for the next decade, is how to combine formal theories and statistical data in a principled and practical way for NLP.

Reviewing the work of the field as a whole over the last fifty years, and what has been achieved, we find first, that the implications of computation in terms of the need for explicit data detail, proper process specification, and appropriate and adequate formalisms are now understood even if sometimes, as in the discourse area, it is too often taken for granted that outline theories can be translated into viable programs. The enormous improvements in machine technology have also meant, very usefully, that it is less essential than it was to worry about proliferations of alternatives during processing, while at the same time, whatever the attractions of cognitively convincing approaches, NLP can be well done in a purely engineering spirit. Moreover while major systems rest on person-decades of experience and effort, it is now possible, with present computing resources, to 'run-up' surprisingly powerful systems and to conduct impressively large experiments in a matter of months or even weeks.

In terms of what language processing requires, and specifically general-purpose language processing, most progress has been made in the area of syntax, where we have effective means of grammar characterisation and useful techniques like chart parsing. More generally, workers in the field now have a stock of conceptual tools, like typed feature structures or case and domain frames, and enough experience of using them to put together a system or interface subsystem for many experimental or developmental purposes and even, for suitably restricted tasks or limited output expectations, for regular operational production. Performance can nowadays, moreover, be improved by exploiting probabilistic information. Advances in low-level speech processing have meant not only that performance in speech recognition without language understanding (as for dictation) is advancing, but that it is now possible to seek some speech understanding systems with language processing capabilities not too far behind those for systems with typed input.

It is nevertheless the case that the most effective current systems, from the point of view of language understanding, are either those with the most limited domains or those with the least demanding tasks. The former include both systems based on putatively general-purpose machinery, customised in a tidy way, and systems essentially designed for given applications. In either case, though the tasks undertaken are not trivial, the systems operate within narrow bounds, for instance in relation to providing explanatory responses in dialogue, and are in general extremely brittle. Moreover while customising may be easier from a solid all-purpose base, there is so far little evidence for large performance gains for this rather than from the ad hoc approach. Overall, the challenge of taking the necessary step from a focused experiment or even convincing prototype to a full-scale rounded-out NLP system with consistent, high-quality performance has not been overcome. In practice, some commercial MT systems, especially when tailored for applications, probably represent the most solid working systems, but they depend on enormous investment and are far from perfect.

These restrictions and limitations emphasise the need for evaluation in user contexts. This

is not only a problem for exigent tasks like information extraction, but is also one for the less demanding tasks, like document retrieval, where shallow processing may suffice but it is hard to show whether natural performance limits have been reached. Again, while highly modular architectures have been widely accepted, there are still major problems for all but the very limited or most tolerant applications, in determining the distribution of information and effort between the linguistic and non-linguistic elements in a system, and between the general-purpose and domain-specific components. Work on generation also lags behind that on interpretation. Moreover, while NLP workers have enlarged their immediate fields and have begun, in particular, to escape from individual sentences and to handle larger wholes in dialogue and extended text, there are important language-using functions, or tasks, like summarising, that have not been attempted in any truly flexible or powerful way; and there are many linguistic phenomena, including ones as pervasive as metaphor, on which work can be hardly said to have begun. It is also the case that while appropriate forms of reasoning, like abduction, have spread from AI generally into NLP and have found useful application at more than one level of processing, there are still very intractable problems to be overcome in providing the apparatus needed to manipulate beliefs and intentions in supporting language use. It may be possible to build useful systems that rely only on discourse conventions and habits, but these are unlikely to be the really powerful systems that everyone wants.

References

[1] ALPAC: *Language and machines: computers in translation and linguistics*, Report by the Automatic Language Processing Advisory Committee, National Academy of Science, Washington DC, 1966.

[2] Alshawi, H. (ed.) *The Core Language Engine*, Cambridge, MA: MIT Press, 1992.

[3] Bledsoe, W. 'I had a dream: AAAI presidential address, 19 August 1985', *The AI Magazine* 7 (1), 1986, 57-61.

[4] Bobrow, D.G. and Collins, A. (eds.) *Representation and understanding*, New York: Academic, 1975.

[5] Booth, A.D. (ed.) *Machine translation*, Amsterdam: North-Holland, 1967.

[6] Brady, M. and Berwick, R.C. (eds.) *Computational models of discourse*, Cambridge, MA: MIT Press, 1983.

[7] Briscoe, E. et al. 'A formalism and environment for the development of a large grammar of English', *IJCAI 87: Proceedings of the 10th International Joint Conference on Artificial Intelligence*, 1987, 703-708.

[8] British National Corpus, *BNC users reference guide*, http://info.ox.ac.uk/bnc/, visited October 2001.

[9] Ceccato, S. 'Correlational analysis and mechanical translation', in Booth 1967, 77-135.

[10] Cohen, P.R., Morgan, J. and Pollack, M.E. (eds.) *Intentions in communication*, Cambridge, MA: MIT Press, 1990.

[11] Cullingford, R. 'SAM', 1981; reprinted in Grosz et al. 1986, 627-649.

[12] Fellbaum, C. (ed.) *WordNet: an electronic lexical database and some of its applications*, Cambridge MA: MIT Press, 1999.

[13] Findler, N.V. (ed.) *Associative networks*, New York: Academic, 1979.

[14] Green, B.F. et al. 'BASEBALL: an automatic question answerer', 1961; reprinted in Grosz et al., 1986, 545-549.

[15] Grosz, B.J., Sparck Jones, K. and Webber, B.L. (eds.) *Readings in natural language processing*, Los Altos, CA: Morgan Kaufmann, 1986.

[16] Harris, L.R. 'Experience with INTELLECT', *The AI Magazine* 5(2), 1984, 43-50.

[17] Hays, D.G. *Introduction to computational linguistics*, London: Macdonald, 1967.

[18] Hendrix, G. et al. 'Developing a natural language interface to complex data', 1978; reprinted in Grosz et al. 1986, 563-584.

[19] Hutchins, W.J. *Machine translation*, Chichester, England: Ellis Horwood, 1986.

[20] Hutchins, W.J. (ed.) *Early years in machine translation*, Amsterdam: John Benjamins, 2000.

[21] Joshi, A.K., Webber, B.L. and Sag, I.A. (eds.) *Elements of discourse understanding*, Cambridge: Cambridge University Press, 1981.

[22] Kamp, H. and Reyle, U. *From discourse to logic*, Dordrecht: Kluwer, 1993.

[23] Kittredge, R. and Lehrberger, J. (eds.) *Sublanguage: studies of language in restricted semantic domains*, Berlin; Walter de Gruyter, 1982.

[24] Kobsa, A. and Wahlster, W. (eds.) *User modelling in dialogue systems*, Berlin: Springer-Verlag, 1989.

[25] Lea, W.A. (ed.) *Trends in speech recognition*, Englewood Cliffs, NJ: Prentice-Hall, 1980.

[26] Locke, W.N. and Booth, A.D. (eds.) *Machine translation of languages*, New York: John Wiley, 1955.

[27] Manning, C.D. and Schuetze, H. *Foundations of statistical language processing*, Cambridge, MA: MIT Press, 1999.

[28] McKeown, K.R. *Text generation*, Cambridge: Cambridge University Press, 1985.

[29] Minsky, M. (ed.) *Semantic information processing*, Cambridge, MA: MIT Press, 1968.

[30] Penn Treebank, http://www.cis.upenn.edu/~treebank/, visited October 2001.

[31] Plath, W. 'Multiple path analysis and automatic translation', in Booth 1967, 267-315.

[32] Rustin, R. (ed.) *Natural language processing*, New York: Algorithmics Press, 1973.

[33] Schank, R.C. 'Language and memory', 1980; reprinted in Grosz et al. 1986, 171-191.

[34] Small, S.L., Cottrell, G.W. and Tanenhaus, M.K. (eds.) *Lexical ambiguity resolution*, San Mateo, CA: Morgan Kaufmann, 1988.

[35] Sparck Jones, K. 'Thesaurus', *Encyclopedia of artificial intelligence* (ed Shapiro), 2nd ed, New York: Wiley, 1992, 1605-1613.

[36] Sparck Jones, K. and Galliers, J.R. *Evaluating natural language processing systems*, Berlin: Springer, 1996.

[37] Sparck Jones, K., Gazdar, G., and Needham, R.M. (eds.) *Computers, language and speech: formal theories and statistical data. Philosophical Transactions of the Royal Society, Series A*, Volume 358, Number 1769, 2000, 1225-1431.

[38] Sundheim, B.M. and Chinchor, N.A. 'Survey of the Message Understanding Conferences', *Proceedings of the ARPA Workshop on Human Language Technology*, San Francisco: Morgan Kaufmann, 1993, 56-60.

[39] Wahlster, W. (ed.) *Verbmobil: foundations of speech-to-speech translation*, Berlin: Springer, 2000.

[40] Winograd, T. 'A procedural model of language understanding', 1973; reprinted in Grosz et al. 1986, 249-266.

[41] Woods, W.A. 'Semantics and quantification in natural language question answering', 1978; reprinted in Grosz et al. 1986, 205-248.

[42] Yngve, V.H. 'MT at MIT', in Booth 1967, 451-523.

[43] Young, S.J. and Chase, L.L. 'Speech recognition evaluation: a review of the US CSR and LVCSR programmes', *Computer Speech and Language*, 12, 1998, 263-279.