[From: *Current issues in computational linguistics: in honour of Don Walker*, ed. Antonio Zampolli, Nicoletta Calzolari, Martha Palmer (Linguistica Computazionale, vol. 9-10); Pisa, Dordrecht, [1994]

Natural Language Processing: A Historical Review*

Karen Sparck Jones
Computer Laboratory, University of Cambridge
e-mail: kn11@phx.cam.cl

Abstract

This paper reviews natural language processing (NLP) from the late 1940's 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 by an attack on 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 forty years of effort in the field.

1 Introduction

At the ACL Conference in 1987 Don Walker, Jane Robinson and I were talking about when we began in NLP research. Fred Thompson told us he began in 1954 and others, like Martin Kay, started out too in the fifties. Work in the 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. It is nice to believe that research in NLP, like scientific research in general, advances in a consolidating way, and though there may be more faith than substance in this, we can certainly do NLP now we could not do in the fifties. We may indeed be seduced by the march of computing technology into thinking we have made intellectual advances in understanding how to do NLP, though better technology has also simply eliminated some difficulties we sweated over in earlier years. But more importantly, better technology means that when we return to long-standing problems they are not always so daunting as before.

Those, like Don, who had been around for a long time, can see old ideas reappearing in new guises, like lexicalist approaches to NLP, and MT in particular. But the new costumes are better made, of better materials, as well as more becoming: so research is not so much going round in circles as ascending a spiral, if only a rather flat one. In reviewing the history of NLP, I see four phases, each with their distinctive concerns and styles. Don, in one way or another and like all of us, to some extent moved in

*The material in the earlier part of this paper is taken from my article "Natural language processing: an overview", in W. Bright (ed.) *International encyclopedia of linguistics*, New York: Oxford University Press, 1992, Vol. 3, 53-59.

time to the current beat. But it is noteworthy that in the push he made for linguistic resources, like corpus collections, he not only significantly promoted what I have called the data-bashing decade that is now with us, but also returned to what was a major concern in the first period of NLP research: building the powerful and comprehensive dictionaries that serious NLP applications, like MT, need.

I define the first phase of work in NLP as lasting from the late 1940s to the late 1960s, the second from the late 60s to the late 70s and the third to the late 80s, while we are in a clear fourth phase now.

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).

Though the period ended under the cloud of the 1966 ALPAC Report, (ALPAC, 1966; Hutchins, 1986), 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, 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 knowledge 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 constructing 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 of both 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 representation and the use of world knowledge then became an argument, reflecting a widespread feeling in AI stimulated by Minsky's promulgation of frames (Minsky, 1975), 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, 1975; 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 Nets (Schank, 1975), 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 tasks 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 signifi-

cantly 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).

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 work on logics for meaning and knowledge representation that can deal with the language user's beliefs and intentions, and can capture discourse features and functions like emphasis and theme, as well as indicate semantic case roles. The issues in this approach are those both of reflecting the refinements of linguistic expressions in indicating time and mood or conveying presuppositions, and of preserving cohesive and coherent discourse structure. However the belief that the grammatico-logical route is the right, because principled, way to go did lead by the end of the 80s to the development of powerful, general-purpose processors, of which SRI's Core Language Engine (Alshawi, 1992) can be taken as an exemplar. These processors could be used to support application systems with at least as much operational power as ones based on less absolutist views, for example on transition nets and frames, and with more potential for superior performance when challenged.

The grammatico-logical approach was also influential in some other ways. It led to the widespread use of predicate calculus-style meaning representations, even where the processes delivering these were more informal than the purist would wish. It also led, when taken with the challenge of building effective systems for, e.g., database query, to a shift in the meaning of "semantic" and "pragmatic" and to changes in the distribution of effort over the system as a whole. Semantic interpretation, given basic lexical data, concentrated on e.g., quantifier interpretation, and the full meaning of expressions was taken to be supplied by reference to the pragmatic context, subsuming both the prior discourse context *and* the application's domain or world model.

All together, the period can be seen as one of growing confidence and consolidation, partly encouraged by the general enthusiasm associated with the Fifth Generation enterprise, but also well-justified by the ability to build better systems, itself reflected in the beginning of the ACL's series of Applied NLP Conferences.

In relation to the central concerns of NLP, consolidation is most evident in syntax, the area in which, from an historical point of view, most progress has been made. By the end of the 1980s, practical system builders could take advantage of relatively well-understood forms of grammar and parsing algorithm, and also sometimes of large actual grammars and bodies of software, like those of the Alvey Natural Language Tools (cf. Briscoe et al., 1987). At the same time, other operational systems joined SYSTRAN and METEO (cf. Hutchins and Somers, 1992) in NLP applications, which now addressed a range of tasks including, e.g., message processing as well as translation, and commercial systems were both offered and purchased, especially for database query (cf. Engelien and McBryde, 1991). Research and development extended world-wide, notably in Europe and Japan, aimed not only at interface subsystems but at autonomous NLP systems, as for message processing or translation. However there was to some extent a division in this period between those focusing on principles and those focusing on practical applications, who did not always follow the formalist, grammatic-logicist

line but exploited whatever conceptual apparatus was to hand, like case and domain frames.

The revival of MT was a significant feature of this period, in which European and Japanese interest played a major part. The European Commission both used production systems based on customised pragmatism and promoted the Eurotra research project on multi-lingual translation within a common, well-defined transfer framework. There were several active Japanese teams, with some translation products in the market (Nagao, 1989). Much of the MT work done assumed that something at least useful and perhaps more could be provided, particularly for specific applications, with or without editor or user participation in the translation process; and it reflected the current state of NLP in grammar choices and the use of modular system architectures.

On the research side, the period was notable for a growth of interest in discourse, and it saw the first serious work on generation, especially multi-sentence text generation. There were two sides to the interest in discourse, which came together in the context of interactive, dialogue systems, for instance for advice giving, where the need for cooperative system responses implies modelling of the participants' beliefs, goals and plans, and can naturally lead to the production of paragraph-length output, for instance in providing explanations. Work on user modelling, as illustrated in Kobsa and Wahlster (1989), was one strand in research on language use intended for active communicative purposes and on discourse structure as related to such purposes (Cohen et al., 1990). At the same time, as e.g., McKeown (1985) showed, rhetorical schemas could be used as convenient recipes for producing communicatively effective, as well as linguistically coherent, text.

From the point of view of NLP as a whole on the other hand, there was more novelty in the connectionist approaches explored in this period, implying a very different system architecture from the conventional modular one (cf. Rumelhart et al., 1986). This work, though not directly absorbed into the mainstream, can be seen as one source, via the idea of probabilistic networks, for the present interest in statistically-flavoured NLP.

The final trend of the 80s was a marked growth of work on the lexicon. This was stimulated by the important role the lexicon plays in the grammatico-logical approach and by the needs of multi-lingual MT, and also by the problems of transportability, customising and knowledge acquisition in relation to individual applications. The first serious attempts were now made to exploit commercial dictionaries in machine-readable form, and this in turn led to the exploitation of text corpora to validate, enhance or customise initial lexical data, research made much easier by the rapidly increasing supply of text material. This last trend can be seen now to be giving the current fourth period of NLP its dominant colour.

5 Phase 4: Late 1980s Onward

Thus the last few years have seen a conspicuous move into statistical language data processing, so much so that this phase can perhaps be labelled the massive data-bashing period. Work on the lexicon has in part concentrated on the development of suitable general formalisms for expressing lexical information, closely tied to the way this is applied through operations on feature systems in syntactic and semantic processing, and taking advantage of AI experience in knowledge representation by viewing the lexicon

as a terminological knowledge base. But this work has been supported by notable initiatives in data gathering and encoding, and has encouraged a surge of interest in the use of corpora to identify linguistic occurrence and cooccurrence patterns that can be applied in syntactic and semantic preference computation. Probabilistic approaches are indeed spreading throughout NLP, in part stimulated by their demonstrated utility in speech processing and hence sometimes advocated not just as supports, but as substitutes, for model-based processing.

The rapid growth in the supply of machine-readable text has not only supplied NLP researchers with a source of data and a testbed for e.g., parsers. The flood of material has increased consumers' pressure for the means of finding their way round in it, and has led both to a new focus of NLP research and development in message processing, and to a surge of effort in the wider area of text processing which deals with the identification of the key concepts in a full text, for instance for use in text retrieval (cf. Jacobs, 1992). Thus NLP, earlier not found to be sufficiently useful for document retrieval based on abstracts, may contribute effectively to searching full text files. All of this work has encouraged the use of probabilistic tagging, originally applied only in data gathering, and the development of shallow or robust analysers. In this context, NLP workers have also been forced to handle more than well-formed individual sentences or well-mannered ellipses and to deal, for instance, with the variety of proper names.

The interest in text, as well as in improving the scope and quality of interfaces, has also promoted work on discourse structure, currently notable for the interaction between those approaching the determination and use of discourse structure from the point of view of computational needs and constraints, and those working within the context of linguistics or psycholinguistics.

A further major present trend can be seen as a natural outcome of the interaction between consumer (and funder) pressures and the real as well as claimed advances in NLP competence and performance made during the 1980s. This is the growth of serious evaluation activities, driven primarily by the (D)ARPA conferences (cf. HLT, 1993) but also reflecting a wider recognition that rigorous evaluation is both required and feasible when systems are solid enough to be used for non-trivial tasks (Galliers and Sparck Jones, 1993). Designing and applying evaluation methodologies has been a salutary experience, but the field has gained enormously from this, as much from learning about evaluation in itself as from the actual, and rising, levels of performance displayed. However evaluation has to some extent become a new orthodoxy, and it is important it should not turn into an ultimately damaging tuning to demonstrate prowess in some particular case, as opposed to improving the scientific quality of work in the field and promoting community synergy.

These evaluation initiatives have nevertheless focused attention on the challenge of NLP tasks involving operations on a large scale, like text retrieval from terabytes of material, and the nature of the specific tasks chosen has also had a stimulating effect in cutting across established boundaries, for instance by linking NLP and information retrieval. More importantly, the (D)ARPA conferences have helped to bring speech and language processing together, with new benefits for NLP from the improvements in speech processing technology since the SUR programme of the 1970s. These improvements are indeed more generally promoting a new wave of spoken language system applications, including ones involving translation, already demonstrated for limited do-

main inquiry systems and proposed in a much more ambitious form in the Verbmobil project (Kay et al., 1991; Siemens, 1991).

Finally, this period has seen a significant, new interest in multi-modal, or multi-media, systems. This is in part a natural response to the opportunities offered by modern computing technology, and in part an attempt to satisfy human needs and skills in information management. But whether combining language with other modes or media, like graphics, actually simplifies or complicates language processing is an open question.

6 Where We Are Now

Reviewing developments in the field as a whole over the last forty 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 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 look for 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 has not been overcome. Nagao's (1989) illustrations

of comparable translation performance for different systems is a salutary reminder of how far NLP has to go.

Nagao's examples, however, by showing how different translations may be equally acceptable, also emphasise the need for evaluation in user contexts, which is the key problem 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. 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.

The present phase of NLP work is interesting, however, not only because of the extent to which it demonstrates that some progress has been made since the 1950s, though far less than was then expected or at least hoped for. Some of its characteristic concerns were also those of the 50s: thus as I said at the beginning, NLP has returned to some of its early themes, and by a path on an ascending spiral rather than in a closed circle, even if the ascent is slow and uneven. The present emphasis on the lexicon and on statistical information, as well as the revival of interest in MT and in retrieval, reflect the pattern illustrated, on the one hand, by Reifler's heroic efforts with the Chinese lexicon and translation (Reifler, 1967), and on the other by the earlier semantic classification work reviewed in Sparck Jones (1992). The present phase, like the first one but unlike some intervening ones, also allows for the rich idiosyncracy of language as well as for its stripped universals, and has again shifted the balance between linguistic and non-linguistic resources in language processing towards the linguistic side.

As I noted too, this return to concerns of the first phase of NLP is also a reminder of Don Walker's long-standing interests. While the Mitre work on syntax with which he was concerned (Zwicky et al., 1965) can be seen as contributing to the ample stream of computational grammar research, the concern with text data with which Don's name has been so closely associated in recent years had its foreshadowing in the title of another of his early papers: "SAFARI: an online text-processing system", a title truly symbolic for both Don and the field (Walker, 1967).

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; see also Hutchins (1986), Chapter 8.

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

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

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

[11] Engelien, B. and McBryde, R. *Natural language markets: commercial strategies*, Ovum Ltd, 7 Rathbone Street, London, 1991.

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

[13] Galliers, J.R. and Sparck Jones., K. *Evaluating natural language processing systems*, Technical Report 291, Computer Laboratory, University of Cambridge, 1993.

[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., Sacerdoti, E., Sagalowicz, D., Slocum, J., "Developing a Natural Language Interface to Complex Data", *ACM Transactions on Database Systems*, Vol 3, No. 3, pp 105-147, 1978.

[19] HLT: *Proceedings of the ARPA Workshop on Human Language Technology,* March 1993; San Mateo, CA: Morgan Kaufmann, in press.

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

[21] Hutchins, W.J. and Somers, H.L. *An introduction to machine translation,* London: Academic Press, 1992.

[22] Jacobs, PS. (ed) *Text-based intelligent systems,* Hillsdale, NJ: Lawrence Erlbaum Associates, 1992.

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

[24] Kay, M., Gawron, J.M. and Norvig, P. *Verbmobil: a translation system for face-to-face dialogue,* CSLI, Stanford University, 1991.

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

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

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

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

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

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

[31] Minsky, M., "A framework for representing knowledge," (ed Winston, P.), *The psychology of computer vision,* McGraw-Hill, 1975.

[32] Nagao, M. (ed) *A Japanese view of machine translation in light of the considerations and recommendations reported by ALPAC, USA,* Japan Electronic Industry Development Association, 1989.

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

[34] Reifler, E. "Chinese-English machine translation, its lexicographic and linguistic problems", in Booth 1967, 317-428.

[35] Rumelhart, D.E., McClelland, J.L. and the POP Research Group, *Parallel distributed processing, 2* vols, Cambridge, MA: MIT Press, 1986.

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

[37] Schank, R. C., *Conceptual Information Processing*, Amsterdam, North Holland, 1975.

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

[39] Siemens AG (ed) *Verbmobil: Mobiles Dolmetschgerät; Studie*, Siemens AG, München, 1991.

[40] Sparck Jones, K. “Natural language processing: an overview”, *International encyclopedia of linguistics* (ed W. Bright), New York: Oxford University Press, 1992, Vol. 3, 53-59.

[41] Sparck Jones, K. “Thesaurus”, *Encyclopedia of artificial intelligence* (ed Shapiro), 2nded, New York: Wiley, 1992, 1605-1613.

[42] Walker, D.E. “SAFARI: an on-line text-processing system”, *Proceedings of the American Documentation Institute Annual Meeting*, 1967, 144-147.

[43] Winograd, T. “A procedural model of language understanding”, 1973; reprinted in Grosz et al. 1986, 249-266.

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

[45] Yngve, V.H. “MT at MIT”, in Booth 1967, 451-523.

[46] Zwicky, A.M. et al., “The MITRE syntactic analysis procedure for transformational grammars”, *Proceedings of the Fall Joint Computer Conference, 1965; AFIPS Conference Proceedings* Vol. 27, Part 1, 1965, 317-326.