Introduction: combining formal theories and statistical data in natural language processing

By Karen Sparck Jones[1], Gerald Gazdar[2] and Roger Needham[3]

[1] *Computer Laboratory, University of Cambridge, New Museums Site, Pembroke Street, Cambridge CB2 3QG, UK*

[2] *School of Cognitive and Computing Science, University of Sussex, Falmer, Brighton BN1 9QH, UK*

[3] *Microsoft Research Limited, St George House, 1 Guildhall Street, Cambridge CB2 3NH, UK*

This paper in its final form appears in *Philosophical Transactions of the Royal Society*, Series A, Mathematical, Physical and Engineering Sciences, 1769, 2000, 1227-1238.

The papers in this volume address, or illustrate, the relation between symbolic and numeric approaches to text and speech processing. This is currently an exciting and productive area of research and development in natural language processing research. This introduction summarises the background, lists important questions to be addressed, indicates how the papers relate to these, and draws out major lessons to be learnt from this state of the art collection.

Keywords: formal theory, statistical data, natural language processing

1. Motivation

The question we addressed in this meeting was how best to combine rule-based and statistics-based approaches to natural language processing (hereafter NLP†). More specifically, we thought it would be useful for the text and speech communities to exchange their respective findings and ideas in the light of (i) the growth of corpus-based strategies in text interpretation and generation, until quite recently primarily symbolic and rule-based, and (ii) the interest in enriching speech processing, hitherto predominantly statistics-based, with prior knowledge of a symbolic kind. This is timely given the increasing demand for practical NLP systems able to cope with bulky, changing or untidy material, and the rapid growth of machine resources able to support the demanding data analysis and rule application that this development implies.

The meeting was thus about progress with the paradigm merger adumbrated in Gazdar (1996), between the symbolic and the probabilistic traditions in NLP. In fact

† In this paper, we use 'NLP' in its most general sense to cover both text processing (interpretation and generation) and speech processing (recognition and synthesis) and to embrace theoretical computational linguistics, practical language engineering, and everything in between. On occasions when we need to refer to text or speech processing specifically, then we will use NLP_T and NLP_S to make our intention clear.

there are two aspects to the merger. One, the more important from the theoretical point of view, is the relation between symbolically expressed rules and numerically grounded facts about language. This covers a spectrum from a primarily symbolic account enhanced with statistical information, as when parsing rule preferences are derived from a corpus, to the primarily numeric, as when information-theoretic facts drawn from a corpus are treated as surrogates for symbolic rules. Current developments are about productive ways of moving along, or choosing a specific point on, this spectrum in relation to particular NLP requirements. The second aspect of the merger, as important for practical systems as for stimulating theoretical developments, is the collaboration of the text and speech communities through interaction between their respectively dominant symbolic and numeric approaches.

In what follows we first review the pertinent background and state of the art in NLP. This review leads to a list of questions that need to be answered if symbolic and statistical approaches are to be effectively combined. We then comment on features of the papers and on relations between the text and speech communities, going on to consider some significant broader themes that emerge from the set of papers as a whole. Finally, we suggest some directions for future research.

2. Background

While NLP$_T$ has been subject to changes of fashion, or emphasis, since the 1950s (Sparck Jones 1994), it has been predominantly rule-based: the argument has been, for instance, about the relative contributions of linguistic and world knowledge. This rule basis reflected the joint influence of theoretical linguistics, in revolt against the earlier distributional approach (cf. Pereira's paper†), and the essentially algorithmic nature of computing. From the beginning, however, there have also been those concerned with the evidence of actual language use as opposed to rule-permitted possibility, whether for grammar adaptation to the sublanguage of a domain corpus as in Sager's work (1978), or in explicitly quantitative methods of semantic classification for language and information processing tasks (Sparck Jones 1964/1986). In the ARPA Speech Understanding Research project of the early 1970s, moreover, while most of the teams applied rule-based approaches to the specified inquiry task, the most successful adopted the much simpler strategy of accumulating surface question texts that were likely to be submitted (Lea 1980).

The early 1980s saw the emergence of NLP$_T$ work on the use of corpora, at first relatively informally as a data source but later for more sophisticated modelling, and specifically for probabilistic modelling (cf. Garside, Leech & Sampson 1987; Church 1988). At the same time, by the mid-1980s, NLP$_S$ research focussed on recognition began to show how effective the statistical approach using Markov modelling at both sound segment and word levels could be, a development which has been significantly accelerated in the 1990s by the (D)ARPA speech recognition evaluations (Young & Chase 1998). The idea that this type of method could be applied to much more challenging NLP tasks than word recognition was given rather striking form in the attempt to use it for machine translation (Brown *et al.* 1990).

For the last decade, the actual or potential value of occurrence and co-occurrence information for all language processing components and application tasks has been

† We refer to papers from this volume simply by author names.

recognised, on the one hand in the increasing use of corpus data and on the other by the revival of rather simple forms of grammar, including finite state grammars and transducers, that fit with the surfacey orientation to language processing that the use of frequency data naturally encourages. In the 1990s, the value of statistics-based approaches for some NLP$_T$ tasks, like document retrieval, has been systematically confirmed (Sparck Jones, in press): here the 'rough', predictive task with its central need to handle text in bulk has justified probabilistic models using word incidence data that reflect underlying linguistic relationships. Just as with word recognition in speech processing, the methods applied depend on context-sensitive *learning.*

Within the space of NLP as a whole, however, these are relatively simple tasks or subtasks. The semantic or syntactic categories and relationships of the knowledge sources that they deploy are weak and coarse-grained. Other tasks are more demanding. Research in the last ten years has thus been increasingly concerned with the derivation and application of rules that are more complex in internal structure, invoke finer sets of classes, operate under more constraining conditions, and require richer relationships between the members of a rule set. The growth of this work on so-called 'empirical methods' in NLP$_T$ is well illustrated by a recent special issue of *AI Magazine* (1997). This research has been made possible by the supply of corpus data and, more particularly, by the provision of the annotated 'answer' data that guides supervised learning and, equally importantly, supports the performance assessment that has become an increasingly important element of NLP R&D. The research has also been able to profit from treatments of grammar that have become familiar in computational linguistics, both through the generic use of feature-based approach and by exploiting the lexically-anchored types of grammar illustrated by HPSG, LTAG, and the categorial and dependency grammar variants that have often been employed in NLP$_T$ in the last fifteen years.

The crucial issue for the meeting was therefore, on the one hand, how far an apparatus needed for NLP can be validated by, refined through, or even derived from usage data, and on the other, what type of NLP apparatus, in its forms of rule and process, provides the best framework for this data capture: clearly, accessible data and useful forms interact with one another.

Our focus was thus on the *integration* of explicit theory of language with primitive observation data, not just the *combination* of independent rule-driven and data-driven 'modules' within an overall language processing architecture. The data have, of course, to be interpreted probabilistically. This implies the use of some formal, though not linguistic, data model – information theory, for example. The route to integration is then illustrated in a simple form by the way that the abstract form of data characterisation represented by a Hidden Markov Model can be made more effective, for NLP purposes, by using the features or categories of an independently endorsed, even normatively justified, linguistic theory.

3. Issues

Examining the relation between statistics and rules for NLP in more detail leads to a whole series of questions:

- How are observed data patterns transformed to, or at least connected with, applicable processing rules for actions?
  e.g. *word category dependency* with *output linearisation*, *semantic collocations* and *predication constraints*

- Then, more particularly, what formalisms facilitate this?
  e.g. *dependency transduction models* vs *recursive transition networks*, *Markov decision processes* vs *speech act plans*

- Thus, how are structurally complex conditions to be captured statistically and made as explicit as required?
  e.g. *context-sensitive phrase structure* vs *attribute listing*

- How can rules derived from statistical data be combined to form effective, integrated rule sets, i.e., in the broadest sense, grammars?
  e.g. *phone ngrams* as *pronunciation conditions*, *syntactic category collocations* as *constituent definitions*

- Then again, what formalisms facilitate this?
  e.g. *dependency grammar* vs *categorial grammar*, *segmental phonology* vs *autosegmental phonology*

- Are statistical inputs pertinent to, or available for, all levels of language description?
  e.g. *subcategorization frame* vs *argument structure*, *pragmatic presuppositions* vs *sound sequences*

- Can individual descriptive levels be predominantly characterised by statistical means?
  e.g. *semantic classification* vs *predication structure*

- What ‘unit extent’ can be defined statistically?
  e.g. *dialogue turn* vs *phone digram*, *syntactic phrase* vs *prosodic phrase*

- How finely can ‘unit intent’ be characterised statistically?
  e.g. *lexical sense* vs *syntactic category*, *homonymy* vs *polysemy*

These questions have been presented from the ‘derivational’ perspective. They have obvious ‘modificational’ analogues referring to the situation where there are already some independent rules and the goal is to extend or tune them against usage data. For instance:

- What rule set decomposition styles promote checking against a corpus?
  e.g. *generalisation hierarchy* vs *pattern set*

There are then further groups of questions, first about the learning process(es) that answers to the first set presuppose, second about the procedural and architectural characteristics of NLP systems that combine absolute and probabilistic information, and third about the NLP application tasks in which statistical information has a necessary, or at least important, rather than merely helpful role. Thus in relation to learning we should ask:

- Are there generic learning procedures applicable across a range of linguistic phenomena?
  e.g. *Bayesian classifiers* vs *factor analysis*, *genetic algorithms* vs *inductive logic programming*

- How do we establish adequate sample size, especially for the derivational case?
  e.g. *10M words of text for spelling rules* vs *100M words for word sense capture*

Questions for the structure and operation of (non-trivial) NLP systems include:

- How to relate information about strictly ordered data, e.g. in the speech signal with unordered or only weakly ordered data?
  e.g. *phonetic event transitions* vs *concept repetitions*

- How to merge numerical information, e.g. rule probabilities, when these are derived by distinct methods or from distinct data sources?
  e.g. *word likelihoods* and *concept likelihoods*

- How to combine the information supplied by distinct system components when some are quantified and others not?
  e.g. *acoustic data* vs *logical forms*

- What is the most appropriate basis for system evaluation?
  e.g. *system module token error rates* vs *overall task success metrics*

Finally, for tasks:

- What tasks call for the use of statistical data, and which do not demand it?
  e.g. *document retrieval* vs *translation*

These are hard questions. They are also interdependent, as Pereira notes, for model complexity, generalisation, and sample size. We should not suppose, either, that they can be answered at the level of generality with which they have just been stated. We have laid them out here as a context and guide for interpreting and assessing the specific contributions that the papers in this volume make.

4. The papers

Pereira's paper explicitly provides a starting point for the collection by exploring the past, present and potential future relations between the formal linguistic and the information theoretic traditions. Thus he emphasises the contribution that the richer statistically-grounded models that we can now hope to build could make both to accounts of language and systems for processing it. Rosenfeld and Ostendorf also discuss general issues in enhancing statistical approaches with features derived from formal linguistic theories.

At a more specific level, the papers fall into various groups offering different perspectives on the overall theme. Thus there are papers illustrating the interaction between statistical data and model rules for speech processing, whether in recognition (Carson-Berndsen, Ostendorf, Rosenfeld) or synthesis (McKeown & Pan, Taylor). Others are concerned with text or transcribed speech (Baayen & Schreuder, Gotoh

& Renals). At the same time, the papers address many different language levels from the components of words (Baayen & Schreuder, Carson-Berndsen) through intermediate units like phrases or sentences (Alshawi & Douglas, Gotoh & Renals, McKeown & Pan, Rosenfeld, Taylor), to discourse units like whole dialogue turns (Young), to extended text (Oberlander & Brew), and even to the real world domains that underlie linguistic expressions (Pulman).

Some of the papers start from the use of statistical data and push this past words to capture larger unit regularities and hence higher-level language structure (Alshawi & Douglas, Gotoh & Renals, Oberlander & Brew, Young); others also start from the data but attempt to leverage pattern capture by exploiting independent linguistic features, constraints or rules (Ostendorf, Rosenfeld, also Pulman). But the complementary strategy, starting from the rule end but modifying and developing an initial model in the light of observed usage is also represented (Carson-Berndsen).

Grouping the papers differently, they illustrate a wide range of techniques for capturing statistical regularities and for representing syntagmatic and paradigmatic language structure, in a way suited to linking data and rules, whether working up from the former or top down from the latter: compare, e.g., Alshawi & Douglas with Young, or McKeown & Pan's two strategies. Again, just as the papers attack different language levels, they also address different subtasks within the scope of a comprehensive language processing system, for instance from word recognition in interpretation (Baayen & Schreuder) to style constraints in text generation (Oberlander & Brew). They also illustrate the role of statistically-motivated approaches for some application tasks, like translation (Alshawi & Douglas).

Memory-based techniques, most visibly employed for parsing by Bod (1998), make several appearances: both McKeown & Pan and Taylor employ them for prosodically coherent speech synthesis whilst Baayen & Schreuder propose a system for morphological interpretation in which a parser and a parse memory compete to deliver the most plausible word structure. Memory fails, by definition, for items previously unseen. The 'unknown word' problem is pervasive in NLP and get attention here from Gotoh & Renals, Carson-Berndsen and Baayen & Schreuder.

Finally, there are papers that address what may be called the inputs and outputs for work in this whole area, namely the general requirements for systematically described corpus data as input (Sampson), and (primarily by example) the evaluation of the results of data analysis, both from a methodological point of view and as illustrations of the performance that language processors exploiting statistical resources can currently achieve (McKeown & Pan).

The papers offer many and varied points bearing on the questions asked earlier. But it is also evident that it is far too soon to take any of the questions as definitively answered. From a rather different point of view, however, we can consider the respective contributions that the text and speech communities have made to what have been becoming their shared concerns in NLP, as these are reflected in the meeting's papers.

Thus first, what, in general terms, have NLP_T researchers learned from NLP_S researchers over the last 15 years? Rather a lot, as the following list indicates.

- That systems that embody probabilistic models perform well, often better than purely symbolic systems.

- That there are trade-offs between the subtlety of the models and the sparseness of the data.
- That finite state mechanisms are useful (pace 40 years of linguistics pedagogy).
- That there is no shame in having a large lexicon.
- That no lexicon is ever comprehensive but previously unseen words have to be dealt with.
- That it makes sense to measure performance.

Secondly, what, in general terms, have NLP$_S$ researchers learned from NLP$_T$ researchers over the last 15 years? Not enough to list, it appears. But while this may be because NLP$_T$ has not had much to offer NLP$_S$, it may also be because speech recognition research initially demanded concentration on its own specific transcription need, and this was far more effectively tackled with probabilistic approaches making minimal reference to linguistic notions than might have been expected. And speech synthesis research has, hitherto, found its most pressing problems in phonetics and signal processing. However, as the meeting papers illustrate, recent NLP$_S$ has been importing ideas from outside – primarily from phonology as developed in linguistics and computational linguistics.

NLP$_S$ researchers are now actively seeking more inputs from NLP$_T$, both to push past the performance limits that they come up against in continuous speech recognition and diphone-based speech synthesis and to enable them to build NLP$_S$ task systems. It is therefore fortunate that, as the papers also indicate, the NLP$_T$ community is developing the shallow and data-oriented forms of language description and processor that will fit with those employed in NLP$_S$: part-of-speech tagging is an obvious example.

5. Emergent themes

There are, moreover, some significant common threads in the meeting papers. The first is the emphasis on the central role of the word. The second is the willingness to rewire the canonical circuit diagram for NLP systems, by relaxing accepted divisions of level and unit. And the third is the liberation of NLP$_T$ brought about by the marginalisation of the notion of well-formedness. We consider these themes in the three subsections that follow.

However, as the papers make clear, while the technological confidence is growing – indeed is rampant – the central challenge of getting enough data, of the right sort, remains because we need more of it to extract the more complex information we desire.

(*a*) *Embracing words*

Combining statistics with rules seems to be naturally lexico-centric, a point sharpened by the issue of how to deal with unknown words encountered in discourse. Words, naturally, form a clear link between NLP$_S$ and NLP$_T$, in both cases raising questions about word constitution and, in the interest in phonology evident in some of the NLP$_S$ papers in this volume, again illustrating the interaction between data-based units and model-based units.

NLP$_S$ researchers have always had the word at the centre of their attention – their recognition task is standardly defined as achieving the mapping from the acoustic signal to a sequence of words. But the word has not always been central to NLP$_T$. For much of the history of the field words have been thought of as a kind of necessary evil. What mattered was rules, and specifically rules applying above the level of the word, most importantly at the level of the sentence. For a variety of reasons both theoretical and practical, but all recognising that to build an NLP wall you have to start with bricks – new or old, whole or half, this has changed and the focus of much NLP$_T$ research is now on words. The various syntactic formalisms in wide use among NLP$_T$ researchers are almost all lexicalist in character – thus Alshawi & Douglas employ a variant of dependency grammar, a grammar which is exhausted by its lexicon, so there are no rules in addition to the lexical entries. The widespread use of finite state techniques, including the now ubiquitous *n*-gram approaches imported from NLP$_S$, for such NLP$_T$ tasks as part-of-speech tagging and sense determination, likewise revolve around the word as the central descriptive unit. Even in morphology, where one might expect to see a smaller unit, such as the morpheme, take precedence, the most influential recent approaches have treated the word (or its more abstract cousin, the lexeme) as the fundamental unit, as in Baayen & Schreuder's paper. And, as Pereira points out, the widespread adoption of the word as the central unit of analysis makes it easier to anchor theory in observation: in a conveniently straightforward sense, uses of words are accessible facts.

(*b*) *Crossing borders*

A friend comes round to improve your hi fi. They put the CD player and the amplifier in the bin. They attach your favorite CD directly to a loudspeaker using a couple of crocodile clips. It sounds pretty good. An implausible scenario, perhaps. But, from the perspective of traditional NLP some almost equally implausible things have been going on in field recently.

Linguistics has almost always been packaged as a layer cake with sound (or ASCII) at the bottom, meaning at the top, and a series of neatly differentiated layers in between. Given this cake, traditional NLP has standardly assumed a correspondingly tidy engineering approach to system operation, namely by pipelining. The trick in, say, text interpretation was to get information derived from the input transferred from one point to another along the pipe and compositionally exploited at each.

In machine translation of text, for example, one would work one's way up from the orthographic representation of one language via morphological and syntactic representations to a semantic representation and then down to the orthographic representation of the other language via a different syntax and morphology. But, in Alshawi & Douglas's paper, much of this is simply cut out of the circuit. They map directly from sequences of orthographic words in one language to sequences of orthographic words in another language via syntactic rules of correspondence. There is only one syntactic representation, there is no morphology, and there is no visible semantics.

A traditional 1970s-80s architecture for a speech dialogue system, to take another example, would progress from word recognition through a variety of standard NLP_T interpretation modules to an inference engine and planner with the result passed through NLP_T generation modules and ending up in a text-to-speech box. Now contrast this with what is described in Young's paper. The surgery has been so radical that the patient is no longer recognisably the same person.

Less dramatically, but as McKeown & Pan and Rosenfeld's papers also illustrate, the effect of dealing with naturally occurring data, especially in the speech case, has been to treat the very different sources of information that are pertinent to some specific language interpretation or generation goal as on all fours, for opportunistic use. This is made manageable by the use of features and weak probabilistic, rather than strong rule-based, connections between data items. Thus it does not entail a return to that ill-conditioned beast of the 1970s, a blackboard architecture. And the use of resources derived from data, such as lexical classifications that cut across conventional descriptive types, further facilitates the crossing of levels and the omission of 'standard' components.

There are other ways, along with the dissolution of levels and hierarchy, in which the conventional approaches to NLP are being undermined. Thus as the foregoing implies, the boundaries between segments are getting blurred.

Conventionally, linguistic model building has sought clean boundaries between discourse (and hence grammar) units. But NLP has to work with naturally occurring data, especially speech data, in all its ragged richness. And NLP researchers have now accepted that useful processors can be built that rely only on implicit as opposed to explicit structure characterisation. This has brought with it a recognition that while there are underlying, motivated segments in discourse, they can be treated more 'casually' at the surface without detrimental effects on language processing performance. This is the strategy adopted in Alshawi & Douglas, for instance. There is no requirement that a discourse unit fully satisfy its formal definition. Further, units may be located where quite different descriptive axes, representing distinct forms of discourse annotation, happen to coincide, in a rough but nevertheless useful way. Oberlander & Brew and Young illustrate these points in different ways, the former by an opportunistic imposition of sentence boundaries, the latter by a hospitable view of dialogue turns. Finally, these unit descriptions may be indifferent as to whether one is dealing with linguistic objects or linguistic processes, in the sense in which a probabilistic abstraction like the Hidden Markov Model relates unit descriptions in a way that can be viewed as either static or dynamic.

Thus in adopting a more relaxed attitude to linguistic levels and segments, as mediated by the introduction of probabilities, we get a view of a language system as concerned with the *dynamic* effects of context. This has spread from speech processing (as illustrated by Rosenfeld's paper, for instance) into the treatment of text. In some cases, for some types of information, it may be convenient to view the correlation between elements as static scenery, in others as active movement. This is not a return to the old declarative versus procedural controversy, but rather, from a system point of view, a merging of object and process. The NLP whole cloth, like damask or taffeta, is continually changing in pattern and colour according to view.

(c) Liberating goats

Contact with reality, especially the reality of speech, in NLP has had further subversive consequences.

Computer science and generative linguistics have (at least) three characteristics in common: (i) they put significant effort into distinguishing sheep from goats; (ii) having done so, they put all their remaining effort into the sheep and ignore the goats; (iii) 100% of the sheep have to be treated right – anything less and you've failed. The sheep here are the well-formed expressions of a formal or natural language (be it Java or Javanese) and the goats are strings that are not well-formed expressions of the language in question. The designer of a Java interpreter or compiler spends no time at all trying to get their system to make sense of character strings that look (roughly) like Java, but aren't. Likewise, the linguist developing a formal semantics for Javanese spends no time at all on trying to assign meanings to strings of Javanese words that do not correspond to any legitimate grammatical unit in Javanese.

Speech recognition research does not share these characteristics. Whatever comes in through the microphone[†] has to get treated as a sheep, no matter how goatlike its appearance. But success in dealing with these sheep is a matter of degree. NLP$_T$, on the other hand, is the child of computer science and generative linguistics and thus, for many years, exhibited their common characteristics. There might have been the odd goat lurking in machine readable text but it was always a topic for further research or someone else's problem.

Two factors have led to a change in this state of affairs. Firstly, a wide variety of corpora of naturally occurring text have become available to NLP$_T$ researchers. Secondly, the field has had to get used to competitions in which the performance of NLP$_T$ systems gets evaluated against such corpora. The odd goat does indeed show up and has to be handled. But, more importantly, all kinds of exotic and previously unstudied sheep breeds (Arapawa, Balkhi, Criollo, . . .) turn out to be pervasive in naturally occurring texts. And, when your work is being evaluated against a simple token error rate, what matters most is that which is most common. Thus names for people, places and organizations; time references; dates; currency expressions; previously unseen words[‡]; parentheticals; appositive constructions; and many other objects rarely considered by linguists have found their way onto the NLP$_T$ agenda. Success has had to become a matter of degree.

The impact of this last point has been considerable. It has liberated the NLP$_T$ field theoretically. Instead of simply rejecting an approach a priori on the grounds that it was known that it would fail to detect certain goats[¶], one was at liberty to embrace techniques that had the potential to work well most of the time. Respect for the reality of language use may initially have presented a daunting challenge, as Sampson points out, but it has, by now, stimulated more effective NLP$_T$ processing.

[†] With the marginal exception of environmental noise.

[‡] We anticipate that most of our readers will have encountered three of these earlier in this paragraph.

[¶] As, for example, finite state devices fail on nested relative clauses in the general case, and context free grammars fail on Swiss German verbal complement embedding in the general case.

6. Future directions

All of the papers bring out the fact that the current state of the art in NLP is a work in progress, not a completed work. Each one suggests many lines of further research to pursue – to provide answers to the questions listed earlier, for example. However, in the light of the themes we have identifed, and of the papers taken together, we can identify the following directions for future research:

- Pushing statistical approaches to the limit in different contexts and for different purposes: this would both show where a statistics-based approach is adequate and, when inadequate, help focus on the point where rules and statistics need to be brought together.

- Enhancing rule-based methods with statistical qualifications in an orderly, incremental way: this would help to determine the real fragility of rule-based approaches as well as lead to more comprehensive and hence powerful systems.

- Complementarily, enhancing statistically-based methods with theory-derived features and rules in a controlled way, to determine the added value in processing of formally-motivated devices that are also theoretically justified.

- Investigating word, and hence larger segment, data characterisation schemes, that are hospitable to a wide range of annotation types: this would allow systems to accommodate both rule-based and data-based information more easily.

- Exploring experimental NLP systems with unorthodox architectures: this would support processing strategies open to combinations of rule- and statistics-based methods.

- Developing new types of corpora with properties suited to deeper data extraction studies: this would promote a a wider range of experiments with different forms of rule and data integration.

- Promoting careful technology and task evaluations: this would stimulate experiments designed to assess the impact of new information resources and process organisations on systems performance.

As the reader will discover when they turn to the papers that follow this introduction, NLP is currently a field in ferment, with exciting new ideas and surprising new results constantly emerging.

References

AI Magazine 1997. Special issue on natural language processing. **18(4)**, pp. 13-96.

Bod, R. 1998 *Beyond grammar: an experience-based theory of language.* CSLI Lecture Notes 88. Cambridge: Cambridge University Press.

Brown, P.J. *et al.* 1990 A statistical approach to machine translation. *Computational Linguistics* **16(2)**, 79-85.

Church, K.W. 1988 A stochastic parts program and noun phrase parser for unrestricted text. *Second Conference on Applied Natural Language Processing*, 136-143.

Garside, R., Leech, G. & Sampson, G. (eds.) 1987 *The computational analysis of English*. London: Longman.

Gazdar, G. 1996 Paradigm merger in natural language processing. In *Computing tomorrow* (ed. I. Wand and R. Milner), pp. 88-109. Cambridge: Cambridge University Press.

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

Sager, N. 1978 Natural language information formatting: the automatic conversion of texts to a structured database. In *Advances in Computers* (ed. M. Yovits), Vol 17, pp. 89-162. New York: Academic Press.

Sparck Jones, K. 1964/1986 *Synonymy and semantic classification* (thesis, 1964). Edinburgh: Edinburgh University Press.

Sparck Jones, K. 1994 Natural language processing: a historical review. In *Current issues in computational linguistics: in honour of Don Walker* (ed. A. Zampolli, N. Calzolari and M. Palmer), pp. 3-16. Amsterdam: Kluwer.

Sparck Jones, K. 2000 Further reflections on TREC. *Information Processing and Management* **36(1)**, pp.37-85.

Young, S.J. & Chase, L.L. 1998 Speech recognition evaluation: a review of the US CSR and LVCSR programmes. *Computer Speech and Language* **12(4)**, pp.263-279.