Natural Language Processing: the 90's and Beyond

Ivan A. Sag
Professor of Linguistics and Symbolic Systems
Senior Researcher: Center for the Study of Language and Information (CSLI)
Stanford University

1. Linguistics and NLP: Past and Present

Thirty-five years ago, Noam Chomsky introduced the field of linguistics to new mathematical tools. Post rewrite-systems and other methods developed in the area of the foundations of mathematics provided new and exciting tools that brought mathematical precision to the enterprise of constructing grammars for natural languages. By the mid 1960's, however, Chomsky's doctrine of *competence* versus *performance* had already become firmly established within the field of linguistics. This doctrine legislated that all facts of language use (performance), including all matters of language processing, be ignored in the development of idealized grammars, which were intended to correspond to speakers' abstract knowledge of language (competence). The competence-performance doctrine, still adhered to by the vast majority of linguists in both the US and Japan, has led to a truly unfortunate situation where most so-called 'theoretical' linguists have no interest whatsoever in matters of language processing, whether it be by humans or computers. The bulk of theory construction within linguistics, still based on the 'string rewriting' algebras derived from the early mathematical work imported into linguistics by Chomsky in the 1950's, has systematically ignored processing issues. Hence it is no surprise that the mainstream of grammatical theory, mired in speculations about increasingly baroque representations and 'movement' operations that relate them, has proven to be of little or no value to the international computational community who share the technological dream that someday computers will be able to understand and communicate in natural language.

In the late 1970's, a new movement began within linguistics. Reacting to the complexity of derivational (transformational) analyses of language that dominated the mainstram of linguistics, a new breed of linguist emerged - one who began to interact with colleagues in computer science, artificial intelligence and psychology. Phrase Structure Grammars, which had been dismissed as inadequate for natural languages by Chomsky and others in the early 1960's, were applied with astonishing success to some of the most complex linguistic problems (e.g. unbounded dependency constructions and constraints on coordinate structures) that had remained unsolved by transformationalists working within Chomsky's tradition. In an important paper, Pullum and Gazdar (1982) showed further that the arguments against such approaches (Context-Free Phrase Structure Grammars in particular) that had been advanced by Chomsky and others contained both mathematical and empirical errors. The stage was set for a new kind of linguistics – one that had natural ties to other research within the emerging interdisciplinary field of Cognitive Science.

The early 1980's saw the development of diverse grammatical frameworks that characterized natural languages in terms of constraint satisfaction systems (category structure algebras), rather than derivational systems (string rewriting algebras). These included Categorial Grammar, Lexical- Functional Grammar and Generalized Phrase Structure Grammar (GPSG). As the 1980's unfolded, these frameworks were refined in important ways, leading to the development of Head-Driven Phrase Structure Grammar (HPSG - Pollard and Sag (1987, to appear) and the closely related Japanese Phrase Structure Grammar (JPSG - Gunji (1987)). HPSG is the basis of the current computational work of ou: research group at Stanford (CSLI); JPSG is the basis of the computational linguistics research at ICOT.

2. Constraint-Based Linguistics and NLP

It was an inevitable fact that natural language processing would languish until some sector of the linguistics community broke away from the Chomskyian tradition to develop a conception of language that would support computational experimentation. The problems of language are simply too complex to be solved without the kind of deep understanding of linguistic phenomena that linguists are trained to develop. As noted above, it was only in the late 1970's that this computationally tractable conception of language emerged within linguistics. Since that time, many research groups have emerged in the US, Europe, Korea, Taiwan, Japan and elsewhere where linguists and computer scientists work together to develop computer systems for natural language processing. These systems are of diverse sorts, including database interfaces, machine (or machine-assisted) translation, and expert system interfaces among others.

What are the design properties of the new theories of grammar that make them suitable for computational application? There are at least five such properties that can be isolated.

Partiality. Language processing crucially involves partial information. Communication often takes place in imperfect surroundings where the linguistic signal is incomplete. Language understanding nonetheless proceeds with remarkable accuracy in the face of incomplete information. In addition, word-by-word processing is often suspended mid-sentence, if the overall message is understood on the basis of prior context. One striking example of this is English 'echo questions', as illustrated in (1).

```
(1)   Speaker A:  Mr. Johannes Jakob Schinkenschmecker is coming to dinner
                  tomorrow night.

                                                                 ght
      Speaker B:    o did you say is coming to dinner tomorrow ni
                   h
                  w
                         ||||||||||||||||||||||||||||||||||||||||||
```

Here, speaker A, in all likelihood suspends word-by-word processing of speaker B's utterance somewhere in the indicated region, once (s)he has recognized that the remainder of B's utterance is a repetition of A's own utterance. What examples like this show is that *partial* linguistic knowledge (e.g. the partial linguistic analysis of *who did you*, *who did you say* or *who did you say is*) is deployed in real time.

Unlike derivational theories of grammar, where syntactic knowledge is defined in terms of rules that transform fully specified grammatical representations of sentences (phrase markers) into other such fully specified representations, constraint-based theories characterize all grammatical properties of sentences in terms of identity constraints that linguistic structures must satisfy. A subset of the constraints that hold true of the entire sentence in (1) is true of any substring of that sentence. Thus only a constraint-based theory provides a readily available characterization of the initial substring of an utterance like the one in (1) in terms of partial linguistic description.

Flexibility. Example (1) also illustrates another important aspect of linguistic performance that is embodied in constraint-based, but not transformation-based linguistic descriptions. The partial description provided for an utterance fragment, as noted is simply a set of equations that describe the available grammatical information in logical terms. Grammaticr! equations may be solved in isolation, or such equations may be partially solved by a process that also considers non-linguistic information (represented in a similar logical language) on its way to understanding the interpretational content of an utterance.

As (1) shows, human processors employ grammatical information about a single sentence flexibly, at some points suspending word-by-word application of that information to consult information about prior discourse. But this massive integration of diverse kinds of information in language processing is in no way limited to such exotic types of example. Knowledge of language and knowledge of the world are flexibly and efficiently interleaved in virtually all language processing. Consider, for example, the following examples.

(2) a. The box in the pen went unnoticed for over an hour.

b. The pen in the box went unnoticed for over an hour.

(3) a. She [found [the book] [on the table]] shortly before leaving the library.

b. She [found [the book on the atom]] shortly before leaving the library.

In the process of comprehending (2)a or (2)b, a language user does not wait until all words are recognized and syntactic structure is assigned before selecting the correct sense of the ambiguous word *pen* ('fenced enclosure' in (2)a; 'writing implement' in (2)b). Rather, this disambiguation, which requires consulting (and perhaps reasoning about) knowledge of the world, is completed well before the sentence is completely processed. The examples in (3) illustrate much the same point with respect to ambiguities of syntactic structure (verbal versus nominal modification) that are resolved incrementally on the basis of such facts as *books fit on tables, but not on atoms*. Constraint-based theories of language, in representing linguistic knowledge in terms of logic-based partial descriptions – the same kind of description being developed throughout the world for the representation of and reasoning about world knowledge, allow for an integrated processing regime that consults information of both kinds on an 'as needed' basis.

Monotonicity. Human processors, in the main, also process language accurately. That is, though 'garden pathing' effects can be produced in laboratory situations, it is an important fact about human language processing that incorrect references, parses, or lexical senses are discarded efficiently in real time. A model of such processing must thus in general function monotonically, making intelligent choices at crucial choice points in such a way as to preserve the truth of the partial linguistic information associated with an intermediate stage of the processing of a given sentence. Constraint-based theories of grammar help to make sense of this fact by providing an essentially monotonic characterization of linguistic knowledge which, when coupled with an appropriate theory of how linguistic and non-linguistic knowledge are controlled in real-time, will give rise to an overall processing architecture that is capable of functioning in an essentially monotonic fashion.

Order Invariance. Note in addition that there is no fixed order in which linguistic and knowledge are consulted in language processing. In (4), the verbal inflection serves to disambiguate the word *sheep* (as singular) before world knowledge is consulted to select the 'fenced enclosure' sense of *pen*.

(4) The black sheep that eats in the pen likes daisies.

In (5), on the other hand, world knowledge is consulted first, with the number of *sheep* not being resolved until the plural verb from *eat* is encountered.

(5) The black sheep in the pen always eat daisies.

Unlike derivation-based conceptions of grammar, where transformational rules are required to operate in a fixed order, the equations of constraint-based theories, by there very nature, are order-invariant - if a structure satisfies constraint A and constraint B, it also satisfies constraint B and constraint A. Thus constraint-based linguistic descriptions have a natural compatibility with processing models that may consult the same linguistic knowledge early in one situation, and later in another.

Reversibility. Finally, it is a basic fact about linguistic knowledge that it is used for many different tasks. Comprehension, production, translation, and so forth may all involve very different kinds of processes, yet our knowledge of language in general plays a role in all such processes. Hence a theory of linguistic knowledge should show no bias for one kind of processing situation over another. Derivation-based theories, despite repeated protestations to the contrary, exhibit considerable bias toward production over comprehension. Transformational analyses are known to lead to combinatoric explosion in comprehension (parsing) applications. Constraint-based approaches, by contrast, define a simple set of constraints on the relation between sound and meaning, a set of constraints that can be consulted by any process. They are thus declarative in nature, and conform to the very same standards advocated in the most useful work within the field of logic programming.

Computational linguists throughout the world have recognized the importance of the design considerations just discussed. At ICOT, at CSLI and elsewhere, constraint-based theories of language have been adapted in numerous research projects that have achieved considerable, yet preliminary success in the development of language process technology.

3. The Resolution Problem

Successful as the constraint-based approach to language has been, I will argue in the final section that it must undergo an important revision if we are to address the fundamental problem facing current research in natural language processing (and perhaps AI in general), which I will refer to as the *Resolution Problem*. As is evident from the observations made in the previous section, the system of linguistic constraints that constitute a natural language do not fully determine the interpretation of any given utterance in that language. Linguistic knowledge of diverse kinds is smoothly and effortlessly integrated with world and contextual knowledge by human language users in the process of communicating, yet at present we have only the beginnings of a basic scientific theory of how this integration takes place. The problem then is to provide a basic scientific answer to the following question:

(6)

```
The Resolution Problem for Natural Language Processing

How are diverse kinds of linguistic and non-linguistic
information integrated in real-time natural processing?
```

To see the extent of this problem, which pervades all of language use, consider the following taxonomy of linguistic phenomena that give rise to what we might refer to as *communicative uncertainty*. First there are massive ambiguities in all human languages:

(7)

```
Ambiguity:
  A. Structural Ambiguity
     I forgot how good beer tastes.
       (beer in general, or good beer?)
     I saw the man with the telescope.
       (the man with the telescope, or seeing with the telescope?)
  B. Lexical Ambiguity
     They can build a better pen.
       (writing implement, or fenced enclosure?)
     The robot wouldn't run.
       (wouldn't function at all, or wouldn't move quickly?)
  C. Ambiguity of Scope
     Jones has found a defect in every Toyota with over 100,000 miles.
       (one recurring defect, or a different one for each Toyota?)
     Everyone in the room speaks at least two languages.
       (the same two languages, or possibly different ones?)
  D. Ambiguity of Ellipsis
     Jones likes Smith more than Parker.
       (more than Parker does, or more than Jones likes Parker?)
```

Linguists have succeeded in providing detailed taxonomies of such ambiguities. Furthermore, there now exist constraint-based grammars that provide reasonably successful analyses of the intricacies of these phenomena.

Second, there are many linguistic phenomena that contain *contextual parameters* – essentially pointers to aspects of the context that supply information essential to understanding the message of the utterance:

(8)

```
Uncertainty of Reference
   He is crazy. (Who is he?)
   John is in charge. (John who? in charge of what?)
   She ran home afterwards. (after what?)
   The motor is on. (which motor?)

Uncertainty of Relation
   The nail is in the bowl.
    (nailed into the bowl, or resting inside of it?)
   John's book (the book John owns?/wrote?/edited?)
```

Researchers in linguistic semantics, especially those linguists, logicians, computer scientists and philosophers working within the framework of *Situation Semantics*, have developed a highly successful constraint-based framework for the analysis of such 'parametric' aspects of interpretation.

Third, there are many linguistic phenomena, some bordering on what is traditionally classified as 'metaphor', where the literal interpretation of an expression is made vivid in diverse ways, depending on context, or else 'coerced' into a related interpretation:

(9)

```
Vivification  (general meanings narrowed in context):
   Craig cut the lawn/hair/cocaine/record/rookie.
     (what kind of cutting?)
   Coffee? (The rising intonation conveys either:
     'I am tentatively suggesting an answer to your question',
      e.g. ''What used to be Columbia's most valuable cash crop?;
      or 'I'm asking whether you want some coffee';
      or 'I'm asking you whether this is coffee.')

Coercion
   The Boston Office called.
     (The intended interpretation is that someone associated
      with the Boston office called.)
```

These problems have been studied in depth by linguists and researchers in Artificial Intelligence. Examples of this kind have been modelled in a successful, yet preliminary way by the work on *abductive reasoning* conducted by Hobbs, Stickel, Charniak and others.

Finally, researchers in AI and linguistics have also studied and developed models of a fourth type of communicative uncertainty, which we may refer to as 'uncertainty of import':

(10)

```
Uncertainty of Import:
  I thought Jones was a spy.
    ('I was right all along.' vs. 'I was mistaken.')
  Smith has outstanding penmanship.
     (praise in a letter of recommendation for a calligrapher;
     the kiss of death in a recommendation letter for a philosopher)
```

This phenomenon, which can be thought of simply as 'reading between the lines', involves drawing inferences based on recognition of the speaker's plans and goals. Preliminary computational models of just such reasoning have been developed by a number of researchers, including Drs. Perrault, Pollack, Cohen, and Hobbs at SRI International/CSLI.

In short, we now have a reasonably clear picture of the ways in which language gives rise to communicative uncertainty – a picture of the space in which resolution takes place. We also know something about the kinds of reasoning that must be performed in order to model successful communication. The important challenge now facing research in natural language processing is to bring together the various kinds of research that have achieved this understanding in such a way as to take the next important step, to develop a basic scientific theory of the Resolution Problem.

4. The Future

This next step – crucial if the dream of computer communication with human language is to be realized – is not an easy one to take. One reason for this is that it will become crucial to integrate the *logic-based*, *constraint satisfaction* methods that have been so successful in the linguistic and computational research of the last decade with *probabilistic*, *constraint relaxation* methods of the sort developed within research on *connectionist models* ('neural nets') and parallel processing regimes.

To see why this might be the case, consider the simple case of lexical disambiguation in (11) and (12).

(11) The entire store was in disarray. The pencils were unsharpened. The pens were empty.

(12) The entire ranch was in disarray. The barns were unpainted. The pens were empty.

In a context like (11), the 'writing implement' sense of *pen* is correctly selected within a fraction of a second, while in (12), it is the 'fenced enclosure' sense of *pen* that is chosen as the appropriate 'resolution' of the utterance. Intuitively, it is obvious why this should be so. In (11), there are strong associations between the meanings of the words in the prior context (e.g. *store* and *pencil* and the 'writing implement' sense of *pen*. Similar associations exist between the meanings of the words in (12) and the 'fenced enclosure' sense of *pen*.

If we are to understand the resolution of interpretation that takes place in real-time communication, we must have a way of modelling how these associations among word senses, which intuitively obey some sort of 'analog' regime, function in tandem with more 'symbolic' types of processing, e.g. lexical look-up, the deployment of grammar rules and the application of rules that determine the compositional interpretations of phrases as a function of the interpretations of their parts.

A purely symbolic approach to modelling associative knowledge and processing of this kind would produce both unwanted and unmanageable complexity. Rather, what is called for is a way of recasting the logic-based, symbolic approach to language processing in terms of a more analog conception that will allow both linguistic regularities and varying 'degrees of activation' of particular kinds of information in particular contexts. We must develop a way of representing all linguistic and world knowledge uniformly, so that such representations will fit into a dynamic model of fluctuating activation patterns, associations with contextually salient information, and so forth. Thus only once we have recast our algebraic, 'rigid' constraint satisfaction conception of language as a more fluid, system of 'soft' constraints will we be able to provide more realistic models of language processing that begin to deal with the Resolution Problem. This recasting of the linguistic and computational results of the last decade, integrating the flexible conception of language with well developed methods of probabilistic and associationistic processing from the connectionist paradigm is the most crucial task to be undertaken by the field of natural language processing in the 1990's and beyond.

This is the conclusion that a number of us at CSLI have reached, and the recognition of the importance of this problem has caused us to form a new research group specifically concerned with the development with a theory of resolution in natural language processing. This group currently includes myself as organizer, Prof. Martin Kay (Stanford Linguistics, CSLI and Xerox PARC), Prof. David Rumelhart (Stanford Psychology and Computer Science and CSLI), Dr. Jerry Hobbs (SRI International and CSLI), Prof. Herbert Clark (Stanford Psychology and CSLI), Dr. Martha Pollack (SRI International and CSLI) and Prof. Michael Tanenhaus (University of Rochester Psychology). This group, which represents a diverse set of disciplines and expertise, is perhaps uniquely qualified to approach the crucial research challenge described above.

In my visit to ICOT, I have been fascinated to learn of the recent research and interests of Dr. Koiti Hasida and Prof. Takao Gunji. They have independently isolated the very same research issues described above as crucial for the success of natural language processing in the 1990's and beyond. Recent work by Dr. Hasida in particular has already begun to take the first step in recasting the logic-based conception of language in terms of connectionist models. There is a strong similarity between Hasida's research and ideas currently being discussed, but not yet executed, by Professor Rumelhart and myself, in conjunction with several students at Stanford.

It has become increasingly clear to me during my short visit here at ICOT that an institutionalized joint research effort between American, Japanese, and perhaps other researchers is essential if we are to develop a theory of the Resolution Problem. The joint effort must

Ivan Andrew Sag

Curriculum Vitae

September, 1990

Personal:

Permanent Address: Department of Linguistics. Stanford University. Stanford, CA 94305

Telephone: Office. (415) 723-9017, 725-2323 Home. (415) 858-0240

Date of Birth: November 9, 1949

Citizenship: U.S.A.

Academic Honors and Awards:

1978–1979 Andrew W. Mellon Post-Doctoral Fellowship. Stanford University

1981–1989 Co-Investigator: Sloan Foundation Grants. Stanford Cognitive Science Group

1981–1983 Co-Principal Investigator (with T. Wasow) National Science Foundation Grant (BNS-8102406): *Phrase Structure Grammar and Natural Language* [Renewed 1982]

1982 Dean's Teaching Award. School of Humanities and Sciences Stanford University

1983–1985 Co-Principal Investigator (with T. Wasow) National Science Foundation Grant (BNS-8309780): *Generalized Phrase Structure Grammar*

1983–present Senior Research Scientist. Center for the Study of Language and Information, Stanford University Gift from The System Development Foundation: *Situated Language Project*

1984 Principal Investigator National Science Foundation Grant (DCR84-11612): *International Workshop on Language Generation*

1985–1987 Co-Principal Investigator (with T. Wasow). National Science Foundation Grant (BNS-8511687): *Head-Driven Phrase Structure Grammar.*

1985–1986 Co-Recipient (with T. Wasow). Equipment Grant from The Hewlett Packard Company.

1987 Principal Investigator. National Science Foundation Grant (BNS-8709887): *Workshop and Conference on Linguistic, Computational, and Philosophical Aspects of Generic, Habitual and Mass Terms.*

1987–1988 *Ameritech* Fellowship. University of Chicago.

Teaching Experience:

1988–present Stanford University: Professor of Linguistics

1984–1988 Stanford University: Associate Professor of Linguistics

1979–1984 Stanford University: Assistant Professor of Linguistics

1977–1979 University of Pennsylvania: Assistant Professor of Linguistics

1976–1977 University of Pennsylvania: Visiting Assistant Professor of Linguistics

1979 University of Texas at Austin: Visiting Assistant Professor (Summer)

1980 University of California at Berkeley: Visiting Assistant Professor of Linguistics (Spring)

1982 University of Maryland: Visiting Assistant Professor, Linguistic Society of America's Linguistic Institute (Summer)

1983 UCLA: Visiting Assistant Professor, Linguistic Society of America's Linguistic Institute (Summer)

1984 Harvard University: Visiting Lecturer (December)

1987–1988 University of Chicago: Visiting Professor of Linguistics and Computer Science

1989 University of Rochester: Visiting Professor (Autumn)

1991 University of California at Santa Cruz: Linguistic Society of America's Linguistic Institute (Summer)

Education:

1973–1976 Ph. D. in Linguistics. Massachusetts Institute of Technology. Cambridge, Massachusetts. Dissertation title: *Deletion and Logical Form.* (Supervisor: Noam Chomsky)

1971–1973 M.A. in Linguistics. University of Pennsylvania. Philadelphia, PA.

1967–1971 B.A. (*magna cum laude*), University of Rochester. Rochester, N.Y. Major: Linguistics. Minor: Sanskrit.

1974 Linguistic Institute. University of Massachusetts at Amherst. (Summer)

1973 Linguistic Institute. University of Michigan. Ann Arbor, MI. (Summer)

1972 Linguistic Institute. University of North Carolina. Chapel Hill, NC. (Summer)