NATURAL LANGUAGE PROCESSING

Julia Hirschberg, Bruce W. Ballard, and Donald Hindle

Julia Hirschberg *and* ***Donald Hindle*** *are members of the technical staff in the Artificial Intelligence Research Department of AT&T Bell Laboratories in Murray Hill, New Jersey.* ***Bruce W. Ballard*** *is a member of the technical staff in the Digital Systems Research Department of Bell Laboratories in Murray Hill. Ms. Hirschberg joined AT&T in 1985 and is working on natural language processing, including intonational meaning, discourse structure, and generating synthetic speech from abstract representations. She has a Ph.D. in computer science from the University of Pennsylvania. Mr. Ballard joined the company in 1984 and is working on domain-independent natural language processing. He has a B.S. in mathematics from Lafayette College, Easton, Pennsylvania,*
(continued on page 57)

Researchers in natural language processing (NLP) apply computational techniques to the understanding and generation of human language. Currently, we are focusing on the development of grammatical formalisms and parsing algorithms, appropriate semantic representations for word and sentence meaning, and ways of specifying more elusive meanings that depend on knowledge of the context of utterance. Applications of natural language research include interfaces to expert systems and database query systems, machine translation, text generation, story understanding, and computer-aided instruction. In this paper, we introduce NLP research in general and describe three NLP projects under way at AT&T Bell Laboratories in Murray Hill, New Jersey.

Introduction

Natural language processing (NLP) is the area of *artificial intelligence* (AI) concerned with the automatic generation and understanding of human languages. Theoretical issues in NLP are commonly divided into three areas:

- *Syntax*, the study of sentence structure
- *Semantics*, the study of context-independent meaning
- *Pragmatics*, or *discourse*, the study of context-dependent meaning.

Students of syntax study how the sentences of a language can be composed from smaller units—words and phrases—and how the constraints on this process can be expressed and represented. From a practical point of view, understanding the constraints speakers of English (for example) must obey when generating a sentence such as

The cat is on the mat. (1)

allows us to build systems that will *not* generate *Is cat the mat on.*

and will not consider such possibilities when attempting to recognize English input.

Students of semantics study how to associate meanings with these structures, mapping between words and phrases and the real world or models of it. This allows us to interpret sentences or to generate them by specifying the conditions under which they are true (i.e., their *truth conditions*).

Students of pragmatics/discourse examine how to interpret sentences in the larger discourse context of who speaker and hearer are, what beliefs they believe each other to hold, what has already been said, and the surroundings in which their exchange occurs.

Syntax. Work in NLP syntax has focused on the construction of *grammars* and the specification of *parsing algorithms*. Grammars are specifications of what constitutes a legal string of a language. A simple phrase structure grammar for the subset of English that generates *the cat is on the mat* is shown in Figure 1. A phrase structure grammar is a set of rewrite rules of the form $A \rightarrow B$ in which a nonterminal A is rewritten to a terminal or another nonterminal B. (Nonterminals generally correspond to phrases; terminals to words.)

With such a grammar (G), a parsing algorithm can determine whether a string of input is accepted by G (i.e., whether that string constitutes a legal sentence of the language for which G is a grammar). In the grammar above,

the cat is on the mat

and

the mat is on the cat

are accepted by the grammar, although

on the mat is the cat

is not. The language accepted by this grammar represents only a tiny fragment of English. Most parsers also assign

Nonterminals				Terminals		
S	→	NP	VP	Det	→	the
NP	→	Det	N	N	→	cat
VP	→	V	PP	N	→	mat
PP	→	Prep	NP	V	→	is
				Prep	→	on

Figure 1. Simple phrase structure grammar. *S* = sentence; *NP* = noun phrase; *VP* = verb phrase; *PP* = prepositional phrase; *N* = noun; *V* = verb; *Prep* = preposition; *Det* = determiner.

structures (or *parse trees*) to legal strings in terms of the grammar being used. (See Figure 2.) Parse trees allow NLP systems to distinguish between the various meanings of sentences. This distinction can be shown by attaching the prepositional phrase under different nodes in the parse tree [directly under the *S* (sentence) node or under the *VP* (verb-phrase) node].

From a theoretical point of view, grammars and parsing strategies are of interest insofar as they reflect aspects of human linguistic abilities. An important debate concerns whether or not parsing (in humans) is deterministic (i.e., whether it can avoid such dead ends for legal input, or whether the parser can "backtrack" to an earlier stage in the parse when it has taken a dead-end search path). The section on "Deterministic Parsing and the Lexicon" discusses this question as well as the relationship between grammar and the lexicon.

The formal properties of grammars and the complexity of parsing algorithms are also of mathematical interest. The coverage of grammars (how much of the language they actually handle) and the efficiency of parsers is of practical concern to designers of real NLP systems.

Semantics. Work in semantics has focused on broad questions such as:

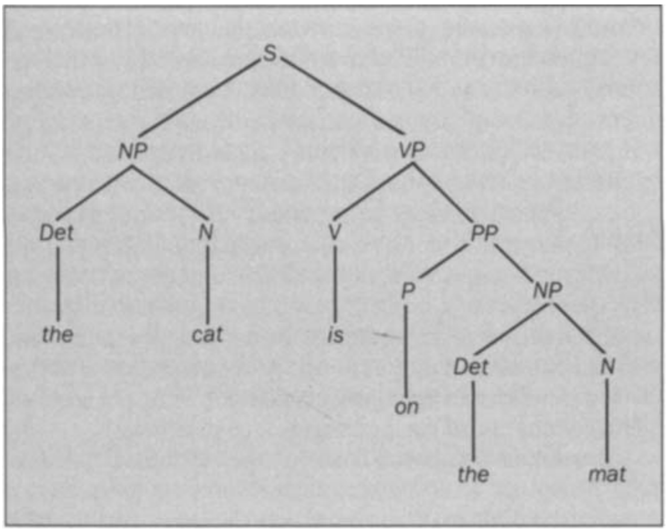

Figure 2. With a parse tree, natural language systems can determine sentence meaning.

- What type of representation is best suited to natural language?
- How do we associate the components of a given representation with real-world counterparts—or some model of these?

The content and structure of a *lexicon* or dictionary is related to the construction of the grammars they access. The study of word meanings, how they are related to one another, and the bases for lexical choice is the subject of *lexical semantics.*

For example, a lexicon should be able to capture the fact that *kill* and *assassinate* describe similar actions but are not interchangeable. A lexicon should also be able to represent the relationship between compounds such as *pet owner* or *polo player* and their subparts, so that not every acceptable variation receives a separate lexical entry but compounds such as *pet player* and *polo owner* will be disallowed.

Semantic form as well as content receives much attention in NLP. Commonly, semantic components of natural language systems map between syntactic structure and the *predicate/argument structure* (or logical form) of a sentence. For example, the predicate/argument structure of Example (1) above might be represented simply as *on(cat, mat).* But whether this mapping should be defined in terms of some "logical" language, such as the first-order predicate calculus, or whether scripts or frames are preferable has been the source of continuing debate. This debate centers around the range of semantic phenomena that must be represented. Among such phenomena, problems of scope ambiguity loom large. For example, sentences such as:

All doors will not open. (2)

John only introduced Mary to Sue. (3)

will have different truth conditions, depending on the scope of *all, not,* and *only.* Example (2) may mean either "no doors will open" or "not all doors will open," depending on whether *not* or *all* has scope over the other. Example (3) may mean

- *John did nothing else with respect to Mary.*
- *John introduced Mary to Sue but to no one else.*
- *John introduced Mary and no one else to Sue.*

depending on the scope of *only.* Other problems include representing natural processes and substances; representing mental attitudes such as beliefs, desires, and intentions; and representing situations, events, and actions.

Pragmatics and Discourse. Although semantic meaning is often identified with *context-independent* meaning, pragmatic meaning is defined as *context-dependent* meaning. However, this division of labor often breaks down. Phenomena commonly accepted as subjects of pragmatic study include *presuppositions,*[1] *conversational implicatures,*[2] *indirect speech acts,*[3] and discourse phenomena such as the various means of communicating *attentional* and *intentional structure* in discourse.[4]

Presuppositions are aspects of utterance meaning that a speaker seems to assume to be true by making the utterance rather than asserting such to be the case. For example, a speaker who utters

The king of France is bald. (4)

might be said to presuppose the existence of a king of France. Hearers appear to treat cases of "presupposition failure" (e.g., the nonexistence of a king of France) differently from cases of false assertion (e.g., the existence of hair on the king of France's head).

Conversational implicatures are another type of non-truth-functional meaning, such as that which can be conveyed by the following:

Some birds fly. (5)

Example (5) may convey that, as far as the speaker knows, it's not the case that all birds fly. However, if we later learn that indeed all birds fly and the speaker knew this, we will not say that the speaker has *lied.* So we will not want to make *not all birds fly* part of the context-independent meaning of *some birds fly.*

Indirect speech acts form yet another type of context-dependent meaning. Utterances such as Example (6):

Can you sit up? (6)

may convey either the direct yes-no question, or a request that a hearer perform some action. (In this case, "sitting up.") Imagine a hospital patient being asked this question by (a) a visitor inquiring about the patient's progress, or (b) a doctor wanting to proceed with an examination. Each of these types of pragmatic meaning has been the subject of investigation in natural language and has been incorporated into natural language question-answering systems.[5,6]

More general studies of discourse or text generation and understanding have focused on identifying discourse structures, scripts,[7] schemas,[8] or rhetorical relations.[9,10] While some of these proposals have permitted text generation in well-defined domains, ability to generalize from domain to domain, dependence on vast amounts of encoded "world knowledge," and difficulties in specifying just what an "elaboration schema" is and how it can be recognized or inferred have plagued all such work.

Recent work by Grosz and Sidner[4] proposes that discourses have three structures, *linguistic, attentional,* and *intentional.* The attentional structure is the structure of what speaker and hearer are attending to during the discourse; it includes the topic structure of the discourse as well as the information structure items mentioned in the discourse whether these items represent "new" or "old" information.

The intentional structure is the structure of what speakers intend to accomplish during the discourse. We can infer both structures, at least partly, from the linguistic structure (the sentences speakers utter). Work by Hirschberg, Pierrehumbert, Litman, and Ward[11,12,13] has proposed some ways intonational features of these utterances can communicate both attentional and intentional structure. (Also, see section on "Intonation and Discourse Structure" below.)

Natural Language Systems. Although NLP systems span a wide variety of applications, the most successful to date have been natural language interfaces to databases.

The number, function, and interaction of the various components of an NL system are important theoretical as well as practical issues. How the syntactic component interacts with the semantic and pragmatic components is particularly important in systems employing deterministic parsers. Proposals for novel system architectures, such as those of connectionist parsers, may also make implicit or explicit claims about the architecture of human linguistic processing. Incorporating knowledge of pragmatics and discourse into more cooperative and user friendly NL systems means that theoreticians as well as system architects must be able to specify how all the information required from context is to be amassed and brought to bear.

Although comparing NL systems is sometimes difficult, evaluation tends to center on issues of *modularity*, *domain-independence*, *machine portability*, *extensibility*, *robustness*, and *grace of degradation*. A system's modularity reflects the extent to which its various components are intelligently partitioned, so that any given module sees all and only the input it needs. A system's domain independence reflects its specialization for a particular application. Both of these are important to a system's *transportability* from one application domain to another. A system that handles quite sophisticated interactions in a very specialized domain may be almost impossible to adapt to another application if the assumptions this interaction rests on must be thought out anew for each application.

How much of transporting can be automated and how much must be done by hand is another measure of transportability. In particular, how much of the lexicon of the new application can be inferred by the system during preliminary interaction with a system administrator or user is an important factor. And portability from machine to machine as well as domain to domain is also a transportability concern. System performance may be measured by how well the grammar extends to new syntactic constructions and the system's ability to handle unforeseen or ill-formed input. Some of these issues are discussed later in the section on TELI, a transportable NL interface.

Deterministic Parsing and the Lexicon

Sentence parsing has two goals. First, a parser is used to discover, in a spoken or written text, meaningful groups of words and their organization, following the grammatical constraints of the language. Second, a parser must provide the units and identify the domains of other processes (i.e., processes such as those that find the predicate-argument structure or the noun-phrase referent of a text). Together, these two parser functions help to limit sentence ambiguity so that the particular meaning of an utterance can be recovered from the vast meaning potential of the language.

Parsing is a process; it concerns the incremental production of the syntactic description of texts. There are two opposing ways to view this process. It can be viewed as a series of choices provided by the grammar that result in a minimal set of alternatives. Or, it can be seen as refining a description, true at each step, to arrive at a useful syntactic description of a sentence. The distinction between these two views is simply whether, in the course of parsing a sentence, alternatives are represented explicitly.

In this section, we describe a research program and present some evidence that parsing should be viewed as a process of incremental description, in which alternatives are never explicitly represented. This approach is known as *deterministic parsing.*[14]

Ambiguity and Overlapping Subsystems. Natural language is particularly blessed with ambiguity. Many words in English appear as different parts of speech depending on context. This is true even of the most common words, such as *to*, which can be a preposition or the auxiliary of an infinitival verb. And even with a single part of speech, many words have multiple meanings. Ambiguity arises from many sources when words are combined in a sentence according to the grammatical rules of the language. Structural ambiguity (the ambiguity that arises from syntax) alone can be staggering. A ten-word sentence typically has several distinct structural analyses, and can have more than one hundred.[15]

Several distinct modules of linguistic knowledge (the phrase structure, predicate-argument structure, noun-phrase reference, and informational structure) contribute to the resolution of ambiguity. No single component has responsibility for resolving all ambiguity. Current research focuses on developing parsers in which each component will say what it can and no more. Each one of these subsystems, (a) phrase structure, (b) predicate-argument structure, (c) reference, and (d) informational structure, has its own organizational principles and makes its own contributions to the description of a sentence.

Look again at the example:

The cat is on the mat.

It is possible to construct an analogous sentence that has the same basic constituents but different lexical material. For example,

My favorite show is on the first Monday of the month. (7)

is another sentence with the structure:

Noun Phrase - Verb - Preposition - Noun Phrase

This sentence means roughly the same as (8) and (9):

The first Monday of the month, my favorite show is on. (8)

My favorite show is on, the first Monday of the month. (9)

The structural difference between (1) and (7) is that in (7), the final noun phrase, *the first Monday of the month,* is not an object of the preposition. Rather it is a time modifier of the entire sentence. And the word *on* is part of an elliptical prepositional phrase that here means something like *on television.*

The problem here is clear. The grammar deals with the organization of phrases and their types, but not with their meaning and reference. But the structural analysis we find for (1) and (7) depends on the meaning and reference of the noun-phrase constituents. How can we capture this fact without explicitly representing the ambiguity?

The solution has two aspects. First, we decompose the notion of *attachment* into a phrase-structure component and a predicate-argument component. In phrase structure, attachment, or *domination,* is interpreted to mean "is somewhere in the phrase." Thus, we can say that the sentence dominates the final noun phrase in both (1) and (7). This is consistent with the prepositional phrase dominating the noun phrase in (1). In predicate-argument structure, attachment means "is an argument of" or "is an immediate constituent of." Both these components must be involved in describing the structure of a sentence.

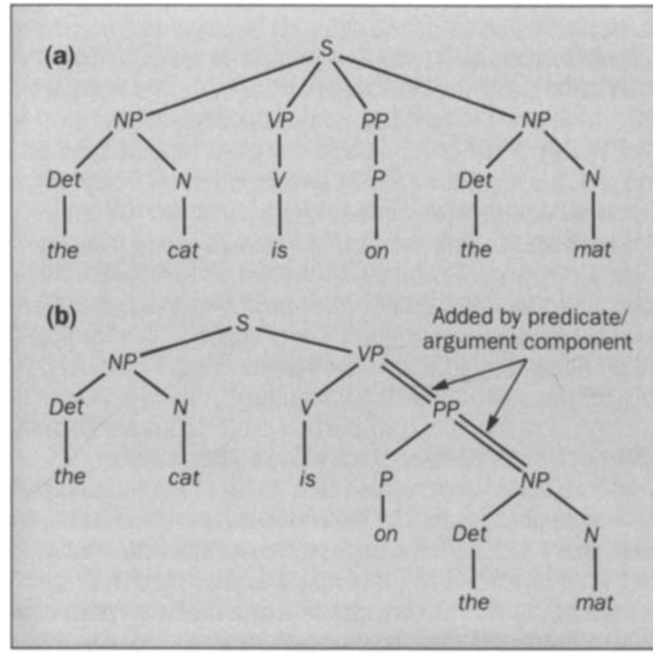

Figure 3. Sample phrase (a) after phrase structure analysis; (b) after predicate-argument analysis. Predicate-argument analysis can go beyond phrase structure analysis and determine whether the noun phrase is the argument of the preposition.

In Example (1), the syntax can find a prepositional phrase *on* and a noun phrase following it. Yet it can remain agnostic about whether the noun phrase is the complement of *on.* The phrase structure analyzer only says what it can be certain of. Essentially, this representational shift succeeds in delaying the decision about the structure by encoding the ambiguity without explicitly representing it.

Thus, the phrase-structure analyses of both (1) and (7) are the same: the phrase-structure analyzer can only say that there is a prepositional phrase *on* that is dom-

Table I. Idiomatic Expressions

Idiomatic Expression	Glossary
On no account	Not for any reason
All in all	Considering everything
As good as	The same thing as
Out of sorts	Grumpy
Lay down one's arms	Stop fighting
Take someone to task	Criticize

inated by the sentence and that following that prepositional phrase there is a noun phrase that is dominated by the sentence. The predicate-argument component has to determine whether the noun phrase is an argument of the preposition. It does this on the basis of having established the referent of the noun phrase, and having looked at the contiguity of the phrases (among other considerations). (See Figure 3.)

The idea of treating parsing as description and using *domination* as a special predicate has wide application. It solves longstanding problems in parsing, including prepositional phrase attachment and conjunction of phrases with *and*. It has been applied successfully to Japanese as well as English.[16,17] In the future, this approach will be systematically applied to other components in the parsing process, in order to identify, for each component, the proper predicates and domains.

The Lexicon and the Grammar. Much of the structure of natural language has never been explicitly described, a rather surprising fact considering that every speaker knows implicitly all the necessary information. This means that finding an architecture for a parsing system is intertwined with the task of finding exactly what information is to be represented. The problem of undescribed linguistic knowledge is particularly striking with respect to the lexicon and its relation to grammar. The lexicon is a vast storehouse of information about the behavior of individual words, but it also contains information about special combinations of words (idiomatic expressions).

Consider the examples shown in Table I. The first four examples are expressions in which all of the elements are fixed. The last two are partially fixed but also contain variables. All these idiomatic phrases conform to the basic phrase-structure rules of English. And all retain, in their idiomatic sense, some of the basic meanings of the constituent words. Thus, idiomatic expressions consist of words put together according to the phrase-structure rules, but additionally endowed with unique features of meaning.

This list of idiomatic expressions could be extended to the tens of thousands. (Estimating the size of the set is difficult, in part at least, because the definition of idiom is not clear.) A dictionary such as the *Oxford Advanced Learner's Dictionary of Current English*[18] with its 50,000 headwords advertises that it includes 11,000 idiomatic expressions. What is so striking about this is that every speaker of English will instantly recognize most of these. Thus, idiomatic expressions represent a vast database of linguistic knowledge, which seems to fit somewhere in between our knowledge of individual words and our knowledge of general phrase-structure rules.

Obviously, knowledge of idiomatic expressions is useful in speaking idiomatic English, but do we really need to know these to process English in general? For at least two reasons, the answer is yes. First, in idiomatic expressions, the meaning is only partially predictable from the parts; general principles can only do part of the job. To find the meaning of an idiomatic expression, it is necessary to know that idiom.

More surprisingly, it turns out that knowledge of these idiomatic expressions is needed in order to decide correctly parts of speech, a crucial step in parsing a sentence. Consider one example, the class of idiomatic expressions consisting of a verb with an infinitive phrase, as in *He had to go*. Examples include *have to, want to, expect to*, and so on. The grammar in general allows the *to* to be either a preposition or a infinitival auxiliary; when the *to* precedes a word that could be a verb, about 5 percent of the instances are prepositional. But, when the *to* precedes

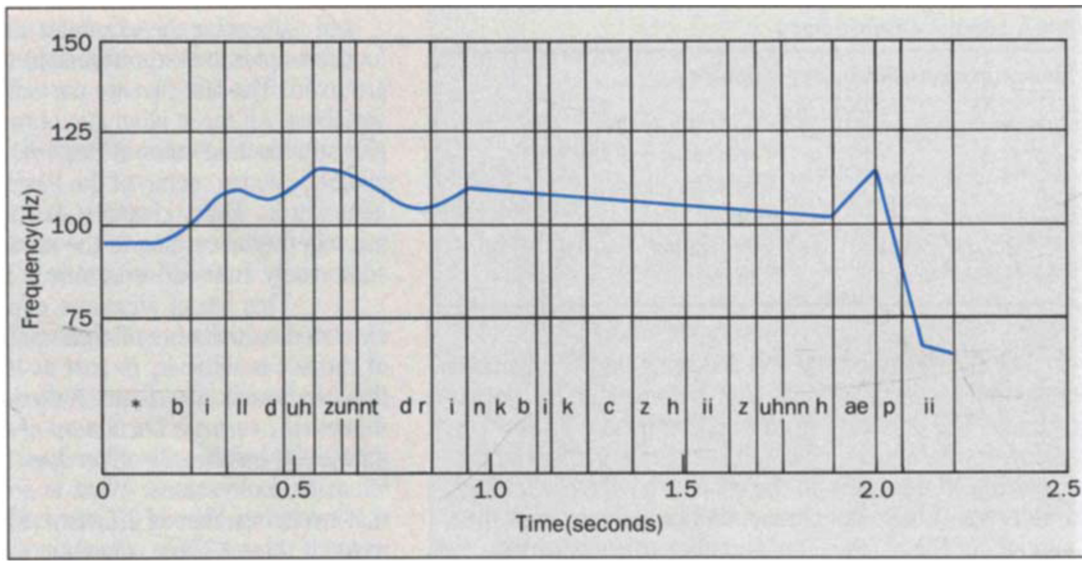

Figure 4. Intonational contours reflect variations in pitch accent, phrase accent, and boundary tone. A declarative contour is shown here.

a word that could be a verb and follows one of these verbs that is in an idiomatic expression, the *to* is always an auxiliary. All 2,654 instances of this construction (i.e., idiomatic verb followed by *to*) in the *Brown Corpus*[19] (a standard million-word sample of published written English) are in fact infinitival expressions. The following sentence from the *Brown Corpus* shows the power idiomatic expressions exert:

> *Also, if we had excluded the ladies*
> *we would have to that extent let the whole world know*
> *at least that much of where we stood.*

For most readers, this sentence requires an extra effort to understand. It is what is called a *garden path* sentence,[14] because the idiomatic expression *have to* misleads the reader into thinking that the *to* is an auxiliary.

How can a parser acquire the large number of idiomatic expressions in a language? We have begun to use existing dictionaries, and indeed these are a rich, largely untapped, source. Nevertheless, existing dictionaries are far from adequate. In addition to the inevitable errors that creep into any dictionary, a more serious problem is omission of information and an inconsistent level of detail. Ultimately, lexical information, including the idioms and fixed phrases that are crucial to syntactic parsing, will have to be acquired automatically. If this is to be successful, we will have to simulate, in part at least, the process that every speaker goes through in learning a language. In this we will make some progress toward answering the general question of how it is that every child acquires the ability to use natural language in a relatively short time and with no formal instruction.

Intonation and Discourse Structure

Recognizing and communicating the *topic structure* of a discourse has long been important in natural language processing. Speakers may generate the same set of sentences but relate them to one another differently, producing quite different messages. How this structural information is conveyed is an important research question.

In this section, we will discuss recent research on

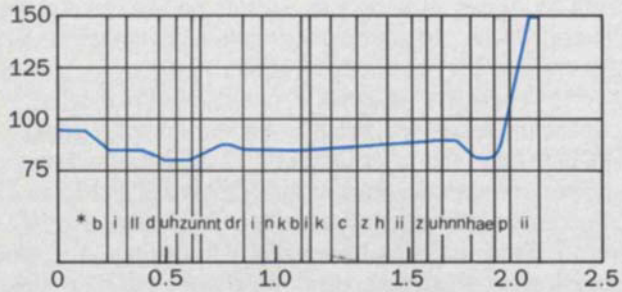

Figure 5. Interrogative contour.

Figure 6. Intonational contour with two phrases.

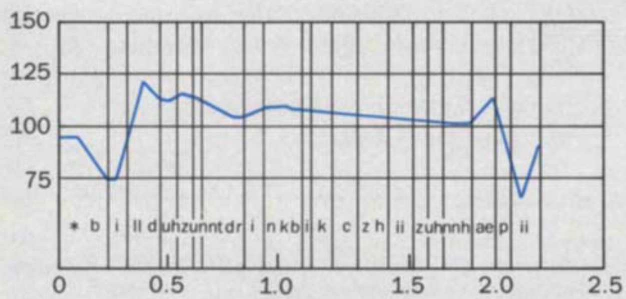

Figure 7. An *L* + H*, or "scooped pitch," accent.

Figure 8. Intonational contour with a larger pitch range.

Figure 9. Intonational contour without final lowering.

the use of intonation to communicate discourse structure. First, however, we will describe the intonational features being examined in terms of the theory of English intonation assumed in this research.

Dimensions of Intonational Variation. In Pierrehumbert's theory of English intonation,[20] intonational contours are described as sequences of *low* (*L*) and *high* (*H*) tones in the f0 (fundamental frequency) contour. A well-formed *intermediate phrase* consists of one or more *pitch accents* (indicated below by "*"), which fall on the stressed syllable of some words, marking items as intonationally prominent, plus an *H* or *L* tone, which characterizes the *phrase accent.* The phrase accent spreads over the material between the last pitch accent of the current intermediate phrase and the beginning of the next (or the end of the utterance). *Intonational phrases* are composed of one or more such intermediate phrases plus a final *boundary tone* at the end of the phrase. The boundary tone may also be *H* or *L* and is indicated below by "%."

Different sequences of pitch accents, phrase accent, and boundary tone define different kinds of *intonational contours*. An ordinary declarative pattern with a final fall is represented as *H* L L%* [that is, a tone with *H** pitch accent(s), an *L* phrase accent, and a *L%* boundary tone]. An interrogative contour is represented as *L* H H%* [*L** pitch accent(s), *H* phrase accent, and *H%* boundary tone]. The contrast between these two contours is illustrated in Figure 4 (declarative) and Figure 5 (interrogative), for the sentence

Bill doesn't drink because he's unhappy.

(These and subsequent intonational contours were synthesized using the Bell Laboratories Text-to-Speech System.)[21]

Intermediate and intonational phrases can be identified from pitch contours by pauses and lengthening of the final syllable in the phrase, as well as the phrase accent and boundary tone. Variation in phrasing is illustrated by comparing Figure 4 (a single phrase) with Figure 6 (two phrases). Phrasing a sentence differently results in differences in meaning. For example, producing *Bill doesn't drink because he's unhappy* as a single phrase conveys that Bill drinks—but for some other reason. As two phrases, it conveys that Bill doesn't drink at all.

There are six types of pitch accent in English,[20] (two simple tones, *H** and *L**, and four complex ones). The high tone, the most frequently used accent, comes out as a peak on the accented syllable (as on *Bill* in Figure 4). *L** accents occur much lower in the pitch range than *H** and are phonetically realized as local f0 minima. The other English accents are composed of two tones. For example, Figure 7 shows a version of the sentence with an *L* + H* accent substituted for the *H** accent on *Bill.* This "scooped" pitch accent is often used to convey uncertainty or incredulity. Figure 7 might convey something like

Maybe Bill doesn't drink because he's unhappy—but X does

or, alternatively,

What do you mean 'Bill doesn't drink because he's unhappy'—of course he does!

depending upon the pitch range used and the speaking rate. Different accent types in general convey different meanings, as do different phrase accents and boundary tones. Together, these produce the more general meanings associated with different intonational contours.

Intonational meaning is also conveyed by variation in overall *pitch range* and *final lowering.* When a speaker's voice is raised, the speaker's pitch range (the distance between the highest point in the f0 contour and the *baseline,* the lowest point a speaker realizes over all utterances) is expanded. Thus, the highest points in the contour become higher and other aspects are affected proportionally. In both cases, the shape of the actual contour is the same, but its scaling is different.

Contrast Figure 4, for example, with Figure 8.

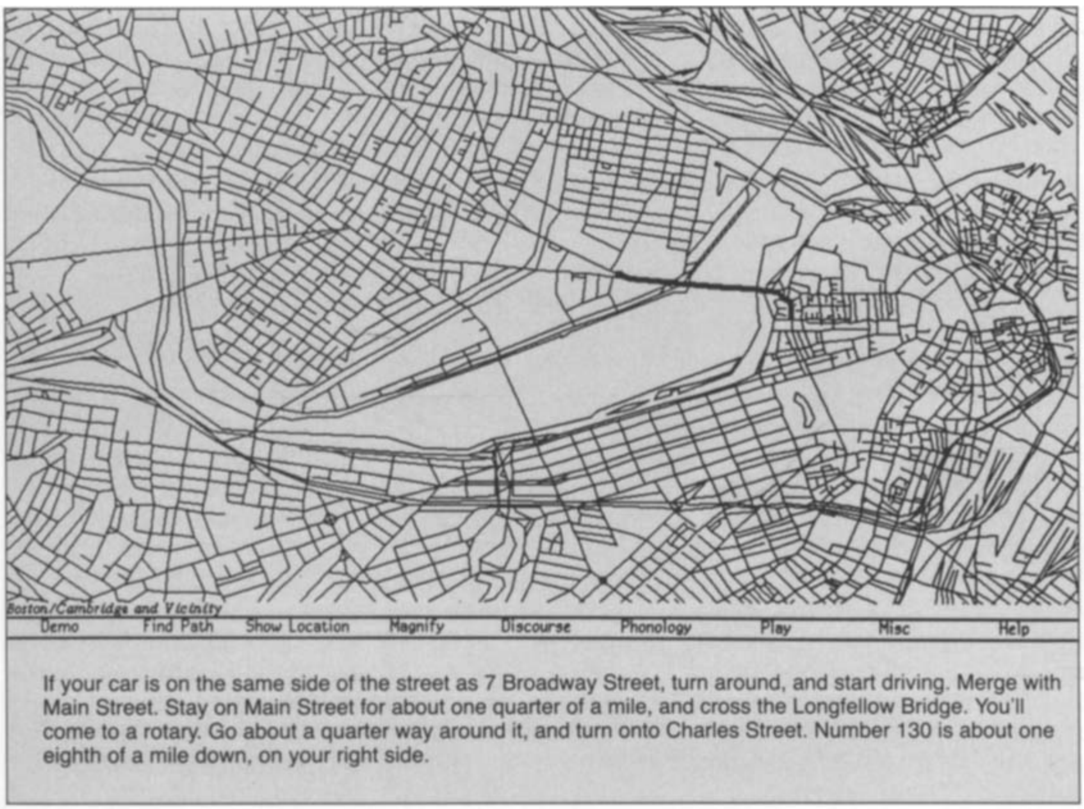

Figure 10. The Direction Assistance system generates maps such as this sample along with a synthetic, spoken directions text that explains the best route to take from one point to another.

The latter depicts the same sentence uttered in a larger pitch range. Variation in pitch range can convey amount of speaker involvement in an utterance as well as the topic structure of a discourse, as discussed below. Another, more local, type of pitch range variation, called *final lowering,* also contributes to the communication of topic structure. Pitch range in declaratives may be lowered and compressed in anticipation of the end of the utterance. This final lowering begins about half a second before the end and gradually increases, reaching its greatest strength right at the end of the utterance.

Contrast Figure 4 (final lowering applied) with Figure 9 (no final lowering applied). While the difference between these two figures may appear visually slight, it is auditorially quite clear. Final lowering appears to reflect the degree to which an utterance is meant to conclude a segment of the discourse; the more final lowering, the more the sense that an utterance "completes" a topic.

Pitch Range and Topic Structure. While it has long been accepted that intonation plays an important role in conveying topic structure, it has not been clear just what this role might be. Recent work by Silverman[22] and by Hirschberg and Pierrehumbert[11] indicate that variation in *pitch range* and *final lowering* conveys discourse structure by marking topic shift and subtopic embedding.

Variation of pitch range and final lowering to con-

Panel 1. Sample text for Direction Assistance

```
T[170] H* If your H* car is on the H* same H* side of the H* street
              as H* ? H* Broadway Street,
        D[60]
    T[153] H* turn H* around,
        D[60]
    T[153] F[.90] and H* start H* driving.
        D[60]
    T[153] F[.90]  H* Merge with H* Main Street.
        D[60]
    T[153] H* Stay on Main Street for about H* one  H* quarter of a H* mile,
        D[60]
    T[153] F[.90] and H* cross the Longfellow H* Bridge.
        D[60]
    T[153] F[.96] You'll H* come to a H* rotary.
        D[40]
        T[137] H* Go about a H* quarter H* way H* around it,
        D[40]
        T[137] F[.90] and H* turn onto  H* Charles Street.
        D[60]
    T[153] H* Number H* 130 is about H* one H* eighth of a H* mile H* down,
        D[40]
        T[137] F[.87] on your L+H* right H* side.
```

vey topic structure has been implemented in an intonation assignment component for *Direction Assistance*, a system that generates synthetic spoken directions for routes in Cambridge, Massachusetts.[23] (See Figure 10.) Direction Assistance was originally developed at Thinking Machines, Inc. The intonation assignment algorithm was designed and implemented at Bell Laboratories during the summer of 1987.

Given a starting point and a destination, Direction Assistance finds the "best" route between the two, using heuristics such as

```
avoid left turns from two-way streets.
```

From this route, the program next constructs a directions text, which is then input to a speech synthesizer. The experimental version currently under development uses the Bell Laboratories Text-to-Speech System.[21] During the construction of this text, the intonational assignment component contributes escape sequences that control intonational variation in the generated speech. The sample text shown in Panel 1 has been generated by the system. Escape sequences are simplified for readability as follows:

- *T* indicates the topline of the current intonational phrase
- *F* indicates the amount of final lowering
- *D* corresponds to the duration of pause between phrases
- *H** or *L** indicates a pitch accent. Other words are not accented.

The topic structure of the text is indicated by indentation.

Note that pitch range, final lowering, and pauses between phrases are manipulated to enforce the desired topic structure of the text. Pitch range is decreased to reflect the beginning of a subtopic; phrases that continue a topic retain the pitch range of the preceding phrase. Final lowering is increased to mark the end of topics. For example, the large amount of final lowering produced on the last phrase conveys the end of the discourse, while lesser amounts of lowering within the text enhance the sense of connection between its parts. Pauses between clauses are also manipulated so that lesser amounts of silence separate pieces of text that are intended to be semantically more related to one another. For example, the segment beginning with

```
You'll come to a rotary
```

is separated from the previous discourse by a pause of 60 centiseconds, but phrases within this segment describing the procedure to follow once in the rotary are separated by pauses of only 40 centiseconds.

Topic structure in Direction Assistance is derived from the structure of the directions themselves, and thus reflects the designers' beliefs about what that structure should be. However, once the route has been constructed, Direction Assistance "knows" the structure it is to convey, and it is a straightforward task to manipulate pitch range, final lowering, and pausal duration to this end.

Systems such as this, which generate speech from abstract representations, thus have a considerable advantage over simple text-to-speech systems. Direction Assistance's intonational component also manipulates accent placement and accent type so that, in general, items already mentioned or otherwise made familiar by preceding discourse are not accented, while "new" items are accented. Accent types are selected according to our current understanding of the meaning that each type contributes to the overall interpretation of the utterance. As our knowledge of intonational meaning increases, we are then able to make modifications to the intonational component.

TELI: A Transportable English-Language Interface

The *Transportable English-Language Interface* (TELI)[24] has been designed to answer English questions about information stored in computerized data retrieval systems. TELI is *transportable* from one application to another; it is also transportable in its ability to adapt to different types of retrieval systems. This contrasts with most natural language question-answering systems, which require that data to be queried be stored as a relational database. TELI is an extension of work done on the LDC (Layered Domain Class) system at Duke University from 1981 through 1984.[25]

TELI is designed to allow a user to ask about any information known to the system and, more importantly, to add or modify information about any new words contained in a particular input.[26] Thus, whenever TELI encounters an unknown word or phrase in an input, it invites the user to specify its part of speech (e.g. "new" is an adjective) and then define its semantics (e.g. a "new" person is one hired within the past year). A variety of customization facilities are available for this purpose.

The semantic sophistication of TELI is such that: (a) definitions of words and phrases may be arbitrarily complex; in particular, they may be any statement in TELI's extended first-order language, and (b) definitions may be given not only by formal means (e.g. responses to system-generated menus) but also by giving an arbitrarily complex English paraphrase or a new word or phrase using language already known to the system. Some of the more complex aspects of natural language semantics dealt with by TELI are negation (*not*, *non*), numerical or logical quantification (e.g., *at least 3*, *all*), and various types of comparatives (*larger than*, *at least 3 more . . . than*).

Sample Interactions with TELI. In the context of the relational database applications to which TELI was first applied, the input

```
Which payroll employees
are not assigned to Boston?
```

produces a relational algebra query that can be roughly paraphrased as

```
select NAME, DEPT, and CITY
    from join PEOPLE-INFO to PROJECT-INFO
    where DEPT = "PAYROLL"
        and CITY != "BOSTON"
```


which produces output such as that shown in Panel 2.

Note that, in addition to providing a literal response (employee name), the system has included auxiliary information that was used in answering the question (department and city). Also note that the required relational `join` operation is generated automatically by the system and is transparent to the user.

Now we will consider in detail the processing of the sample input

```
List the cars made by each US company
that are larger than the Cutlass Ciera.
```

as it filters through each phase of processing. [The car manufacturers' names and model names used here are marks of their respective parent companies.] This time, we will assume that the Kandor knowledge representation system[27] is being used to store the data. (See also the article by Brachman in this issue.)[28]

After each word of the input is looked up in the system dictionary, the input is subjected to syntactic analysis, which yields a parse tree that is then normalized to produce the following normalized parse tree:

```
(CAR (NOUN CAR)
  (COMPAR (ADJ LARGE) >
     (CAR (= CUTLASS-CIERA)))
  (VERBINFO (COMPANY MAKE CAR NIL NIL NIL)
       (OBJ ?)
       (SUBJ (COMPANY (QUANT EACH)
         (COMPANY (ADJ US)
           (NOUN COMPANY)))))
```

In effect, this structure indicates that the user is asking for objects of type `CAR`, as modified by (a) a comparison against the `Cutlass Ciera`, based on the meaning of the adjective `LARGE`, and (b) the meaning of the verb phrase associated with `Companies that make cars` in which the object type being modified corresponds to the

Panel 2. TELI Output for "Payroll employees not in Boston"

Final Output:

Name	Department	City
Jones	Payroll	Baltimore
Smith	Payroll	Buffalo
Taylor	Payroll	Baltimore

direct object of the phrase and where the subject quantifies over `EACH` object of type `COMPANY` to which the adjective `US` applies. Note that the parser has recognized that the verb `made` represents an abbreviated form of the passive phrase `that are made` and thus `CAR` has been appropriately labeled in the parse tree as the object of the verb, not its subject. Note also the attachment of the phrase `larger than the Cutlass Ciera` to the noun `cars`, not the (closer) noun `company`.

Conversion of this expression into a language-independent logical form, which reflects the definitions of the words and phrases found in the input, yields the following logical form:

```
(MAP (λ (P∃)
    (SET (CAR P1)
      (AND (> (LENGTH-OF P1)
              (LENGTH-OF CUTLASS-CIERA))
        (= (COMPANY-OF P1) P∃))))
    (SET (COMPANY P∃)
      (US-COMPANY P∃)))
```

which indicates that a function to find an appropriate set of cars (denoted by the λ form) is to be mapped onto each U.S. company. Note the movement of the scope of the quantifier `each`, which represents one of the types of interpretations imposed by the system.

Taking into account translation information provided by the person who has customized TELI for the backend retrieval system at hand, this logical form is converted, in the context of the Kandor backend mentioned above, into an executable expression:

```
(GPC-FN&ARG ( λ(P∃)
  (SUBSET (λ (P1)
    (AND (> (KSV P1@ S{LENGTH})
      (KSV @I{CUTLASS-CIERA}@ S{LENGTH}))
      (= (KSV P1 @ S{COMPANY}) P∃)))
    (KI @F{CAR})))
  (SUBSET (λ(P∃)
    (KI? P∃ @F{US-COMPANY}))
    (KI @ F{COMPANY})))
```

where @ `I`, @ `F`, and @ `S`, respectively, denote `individuals`, `frames` (concepts), and `slots` that provide the basis of the taxonomic structuring of Kandor; `KI` and `KSV` are retrieval functions associated with Kandor; and the remaining functions capture the semantics of the terms of the logical form language (e.g., `AND` and `SUBSET`).

Finally, this expression is executed, producing labeled lists that are displayed on the computer terminal as shown in Panel 3.

Discussion. Most research in natural language takes the form of either *theoretical* work, which often involves intensive study of one particular aspect of language such as syntax, semantics, or phonetics; or *applied* work, which typically seeks to combine known strategies from each of several areas to achieve a specific goal. In contrast, TELI is an attempt to contribute to the *theory* of *applied* natural language processing. Thus, its main focus has been to develop formalisms and associated algorithms that can interact reliably to provide capabilities that are both effective and adaptable. A comparison of TELI against other fully implemented natural language question-answering systems may be useful. In conclusion, TELI provides:

Panel 3. TELI Output for "US cars larger than the Cutlass Ciera"

```
Final Output:

AMC:          (none)
Buick:        Electra,
              Le-Sabre, and Regal
Cadillac:     De-Ville
Chevrolet:    Caprice and Monte-Carlo
Chrysler:     Fifth-Avenue
Dodge:        Diplomat
Ford:         LTD-Crown-Victoria and
              Thunderbird
Lincoln:      Continental, Mark-VII and
              Town-Car
Mercury:      Cougar and Grand-Marquis
Oldsmobile:   98-Regency, Cutlass-
              Supreme, and Delta-88
Plymouth:     Gran-Fury
Pontiac:      Bonneville and Grand-Prix
```

- Roughly equal coverage of syntax
- Greater-to-much-greater flexibility in the semantics it allows
- Unparalleled customization facilities.

TELI: Current and Future Work. TELI has been designed as a research prototype. Written in Common Lisp, with a modest use of Flavors, it runs on the Symbolics® family of Lisp machines, but should be adaptable to other Common Lisp-Flavors systems.

Some of the current limitations of TELI that should be addressed in the near future are:

- Sentences should be interpreted in the context of previous interactions with the system. This would allow pronouns or other abbreviation devices, including input such as `What about Smith?`
- At least some grammatical errors should be tolerated. Presently, TELI offers spelling correction and some facilities that help a user to track down why an input could not be processed, but allows no grammatical deviance.
- Facilities for spoken input and output are highly desirable. In providing for spoken output, it may be necessary to address nontrivial problems in natural language generation. Specifically, it will be necessary to produce a linear reading of all or some of what is now being presented in tabular form. Future work with TELI will address one or more of these issues.

Conclusion

It is impossible to do justice to such a rich and varied field as NLP in a single overview. For example, we have been forced to omit discussion of the interdisciplinary aspects of the field. We have benefited from collaboration with researchers in philosophy, psychology, and neural science, as well as linguistics. Within AI itself, natural language processing has enjoyed useful interaction with students of planning and reasoning, vision, logic programming, and expert systems. In fact, perhaps the most exciting facets of NLP research are the opportunities for interaction with many disciplines and for immersion in a wide variety of applications problems.

References

1. G. Gazdar, *Pragmatics: Implicature, Presupposition, and Logical Form*, Academic Press, New York, 1979.
2. H. P. Grice, "Logic and Conversation," *Syntax and Semantics*, Academic Press, New York, 1975. From 1967 lectures.
3. J. R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge, England, 1969.
4. B. Grosz and C. Sidner, "Attention, Intentions, and the Structure of Discourse," *Computational Linguistics*, Vol. 12, No. 3, 1986, pp. 175-204.
5. J. Kaplan, "Cooperative Responses from a Portable Natural Language Database Query System," M. Brady, ed., *Computational Models of Discourse*, MIT Press, Cambridge, Massachusetts, 1982.
6. C. R. Perrault and J. F. Allen, "A Plan-based Analysis of Indirect Speech Acts," *American Journal of Computational Linguistics*, Vol. 6, 1980, pp. 167-182.
7. R. Schank, *Conceptual Information Processing*, American Elsevier, New York, 1975.

8. K. McKeown, *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*, Cambridge University Press, Cambridge, England, 1985.
9. J. R. Hobbs and D. Evans, "Conversation as Planned Behavior," *Cognitive Science*, Vol. 4, No. 4, 1980, pp. 349-377.
10. W. C. Mann, "Discourse Structures for Text Generation," *Proceedings of COLING-84*, Association for Computational Linguistics, Stanford, California, July 1984, pp. 367-375.
11. J. Hirschberg and J. Pierrehumbert, "The Intonational Structuring of Discourse," *Proceedings of the 24th Annual Meeting*, Association for Computational Linguistics, New York, 1986, pp. 136-144.
12. J. Hirschberg and D. Litman, "Now Let's Talk About *Now*: Identifying Cue Phrases Intonationally," *Proceedings of the 25th Annual Meeting*, Association for Computational Linguistics, Stanford, California, 1987.
13. J. Hirschberg et al., "Intonation and the Intentional Structure of Discourse," *Proceedings, 10th International Joint Conference on Artificial Intelligence*, Milan, Italy, 1987.
14. M. P. Marcus, *A Theory of Syntactic Recognition for Natural Language*, MIT Press, Cambridge, Massachusetts, 1980.
15. W. A. Martin, K. W. Church, and R. S. Patil, *Preliminary Analysis of a Breadth-first Parsing Algorithm: Theoretical and Experimental Results*, Technical memo, Massachusetts Institute of Technology Laboratory for Computer Science, Cambridge, Massachusetts, 1981.
16. M. P. Marcus, D. Hindle, and M. M. Fleck, "D-theory: Talking about Talking about Trees," *Proceedings of the 21st Meeting*, Association for Computational Linguistics, 1983, pp. 129-136.
17. M. P. Marcus and D. Hindle, "A Computational Account of Extra Categorial Elements in Japanese," *Papers presented at the First SDF Workshop in Japanese Syntax*, System Development Foundation, 1985.
18. *Oxford Advanced Learner's Dictionary of Current English*, 3rd edition, A. S. Hornby, ed., Oxford University Press, Oxford, England, 1974.
19. W. N. Francis and H. Kucera, *Frequency Analysis of English Usage*, Houghton Mifflin, Boston, 1982.
20. J. B. Pierrehumbert, *The Phonology and Phonetics of English Intonation*, Ph.D. thesis, Massachusetts Institute of Technology, September 1980.
21. J. P. Olive and M. Y. Liberman, "Text to Speech—An Overview," *Journal of the Acoustic Society of America, Supplement 1*, Vol. 78, Fall 1985.
22. K. Silverman, *The Structure and Processing of Fundamental Frequency Contours*, Ph.D. thesis, Cambridge University, Cambridge, England, 1987.
23. J. R. Davis, *Giving Directions: A Voice Interface to an Urban Navigation Program*, Media Technology Lab Memo, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1987.
24. B. Ballard, "User Specification of Syntactic Case Frames in TELI, a Transportable, User-customized Natural Language Processor," *Proceedings of COLING-86*, Association for Computational Linguistics, Bonn, Germany, August 1986, pp. 454-460,
25. B. Ballard, J. Lusth, and N. Tinkham, "LDC-1: a Transportable Natural Language Processor for Office Environments," *ACM Transactions on Office Information Systems*, Vol. 2, No. 1, 1984, pp. 1-23.
26. B. Ballard and D. Stumberger, "Semantic Acquisition in TELI: a Transportable, User-customized Natural Language Processor," *Proceedings of the 24th Annual Meeting*, Association for Computational Linguistics, New York, June 1986, pp. 20-29.
27. P. Patel-Schneider, "Small Can Be Beautiful in Knowledge Representation," *Proceedings, IEEE Workshop on Principles of Knowledge-Based Systems*, Denver, Colorado, December 1984.
28. R. Brachman, "Knowledge Representation," *AT&T Technical Journal*, Vol. 67, No.1, January/February 1988, pp. 7-24.

Biographies (continued)

and an A.M. and a Ph.D. in computer science from Duke University. Mr. Hindle joined AT&T in 1985 and is working on developing robust parsing models and comprehensive grammars for analyzing natural language. He has a Ph.D. in linguistics from the University of Pennsylvania.

(Manuscript received October 2, 1987)