Semantic and Discourse Information for Text-to-Speech Intonation

Laurie Hiyakumoto, Scott Prevost & Justine Cassell
The Media Laboratory
Massachusetts Institute of Technology
20 Ames Street, Cambridge, MA USA 02139
{hyaku,prevost,justine}@media.mit.edu

Abstract

Concept-to-Speech (CTS) systems, which aim to synthesize speech from semantic information and discourse context, have succeeded in producing more appropriate and natural-sounding prosody than text-to-speech (TTS) systems, which rely mostly on syntactic and orthographic information. In this paper, we show how recent advances in CTS systems can be used to improve intonation in text reading systems for English. Specifically, following (Prevost, 1995; Prevost, 1996), we show how *information structure* is used by our program to produce intonational patterns with context-appropriate variation in pitch accent type and prominence. Following (Cahn, 1994; Cahn, 1997), we also show how some of the semantic information used by such CTS systems can be drawn from WordNet (Miller et al., 1993), a large-scale semantic lexicon.

1 Introduction

Although theories relating intonational patterns with discourse phenomena have been proposed (Pierrehumbert and Hirschberg, 1990), existing TTS systems, and even CTS systems, often fail to exploit them. The most advanced intonation systems (Hirschberg, 1990; Hirschberg, 1993) have relied on elements of discourse context which are relatively easy to extract from text, such as lexical givenness. Our system augments this approach by analyzing information structure and drawing semantic information from a large-scale semantic database.

Information structure is identified by first dividing utterances into semantic propositions rather than syntactic constituents (cf. (Monaghan, 1994)), in accordance with our belief that intonational domains are often orthogonal to traditional syntactic constituents.[1] These semantic propositions are then sub-divided into *theme* (or *topic*) and *rheme* (or *comment*). The theme of the proposition represents a link to prior utterances, whereas the rheme provides the core contribution—roughly, the new or interesting part—of the proposition. Based on previous intonation generation work (Prevost and Steedman, 1994), thematic and rhematic items requiring accentuation are assigned **L+H*** and **H*** pitch accents respectively.

The notion that large databases in TTS systems can substitute for application-specific knowledge bases has been suggested by (Horne et al., 1993) and (Cahn, 1994). Following Cahn's proposal (Cahn, 1994) and implementation (Cahn, 1997), we employ WordNet to identify lexical items related by synonymy, hypernymy or hyponymy, and also to identify contrastive lexical items. However, our use of information structure situates our work in a different theoretical framework from Cahn's.

[1] (Steedman, 1991) and (Prevost and Steedman, 1994) show how the correspondence between intonational phrasing and semantic constituency can be modeled by Combinatory Categorial Grammar (CCG), a formalism allowing a more flexible notion of syntactic constituency.

In the remainder of this paper we describe how our present TTS research builds on the growing body of CTS research. First we present the motivation for our approach and the underlying theoretical model of intonation. Then we briefly introduce WordNet. Next, we describe the phases of computation and discuss the role of WordNet in making accentability decisions. Finally, we present sample output of the system, explore areas for improvement, and summarize our results.

2 Semantic and Discourse Effects on Intonation

The effects of "givenness" on the accentability of lexical items has been examined in some detail and has led to the development of intonation algorithms for both text-to-speech (Hirschberg, 1990; Hirschberg, 1993; Monaghan, 1991; Terken and Hirschberg, 1994) and concept-to-speech systems (Monaghan, 1994). While the strategy of accenting open-class items on first mention often produces appropriate and natural-sounding intonation in synthesized speech, such algorithms fail to account for certain accentual patterns that occur with some regularity in natural speech, such as items accented to mark an explicit contrast among the salient discourse entities. In addition, the given/new distinction alone does not seem to account for the variation among accent types found in natural speech.[2] Unfortunately, such issues have been difficult to resolve for text-to-speech because of the paucity of semantic and discourse-level information readily available without sophisticated text understanding algorithms and robust knowledge representations.

Previous CTS work (Prevost, 1995; Prevost, 1996; Prevost and Steedman, 1994) showed that both contrastive accentual patterns and limited pitch accent variation could be modeled in a spoken language generation system. The present work incorporates these results in a text-to-speech system, using a similar representation for discourse context (i.e. information structure), and replacing the domain-specific knowledge base with WordNet.

We represent local discourse context using a two-tiered information structure framework. In the higher tier, propositions are divided into theme and rheme. The theme represents what the proposition is about and provides the contextual link to prior utterances. The rheme provides the core contribution of the proposition to the discourse—the material the listener is unlikely to predict from context. In the simplest case, where an utterance conveys a single proposition, the division into theme and rheme is often straightforward, as shown in the question/answer pair in Figure 1.

(Steedman, 1991) and (Prevost and Steedman, 1994) argue that for the class of utterances exemplified by these examples, the rheme of the utterance often occurs with an intonational (intermediate) phrase carrying the **H* L-L%** (**H* L-**) tune, while the theme, when it bears *any* marked intonational features, often carries the **L+H* L-L%** (**L+H* L-**) tune. While this mapping of thematic constituents onto intonational tunes is certainly an oversimplification, it has been quite useful in previous concept-to-speech work. We are currently using the Boston University radio news corpus (Ostendorf, Price, and Shattuck-Hufnagel, 1995) to compile statistics to support our use of this mapping.[3] Preliminary results show that the **H*** accent is most prevalent, occurring more than fifty percent of the time. **!H*** and **L+H*** occur less frequently than **H***, but more than any of the other possible accents. We take the prevalence of **H*** and **L+H*** in the corpus to support our decision to focus on these accent types.

Given the mapping of tunes onto thematic and rhematic phrases, one must still determine which items within those phrases are to be accented. We consider such items to be in theme- or rheme-*focus*, the secondary tier of our in-

[2] Of course, the granularity of the given/new distinction may be at issue here. The relationship of accent types to the given/new taxonomy proposed by (Prince, 1981) may warrant more exploration in a computational framework.

[3] This corpus is partially annotated with ToBI-style (Pitrelli, Beckman, and Hirschberg, 1994) intonation markings.

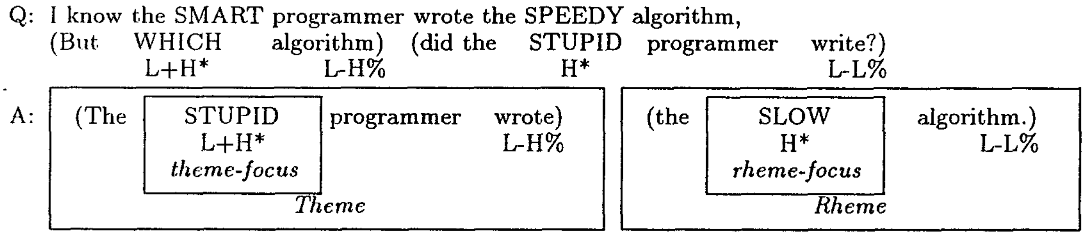
Q: I know the SMART programmer wrote the SPEEDY algorithm,
(But WHICH algorithm) (did the STUPID programmer write?)
L+H* L-H% H* L-L%
A: (The STUPID programmer wrote) (the SLOW algorithm.)
L+H* L-H% H* L-L%
theme-focus
rheme-focus
Theme
Rheme

Figure 1: An Example of Information Structure

formation structure representation, as shown in Figure 1. The determination of focused items is based on both givenness and contrastiveness. For the current TTS task, we consider items to be in focus on first mention and whenever WordNet finds a contrasting item in the current discourse segment. The algorithm for determining the contrast sets is described in Section 4 below.

The adaptation of an information structure approach to the TTS task highlights a number of important issues. First, while it may be convenient to think of the division into theme and rheme in terms of utterances, it may be more appropriate to consider the division in terms of propositions. Complex utterances may contain a number of clauses conveying several propositions and consequently more than one theme/rheme segmentation. Our program annotates thematic and rhematic stretches of text by first trying to locate propositional constituents, as described in Section 4.

Another information structure issue brought to light by the TTS task is that themes may not consist solely of background material, but may also include inferable items, as shown in example (1). In this example, "name" is certainly not part of the shared background between the speaker and the listener. However, since it is common knowledge that pets have names, it serves as a coherent thematic link to the previous utterance.[4]

(1) Miss Smith has a Collie.
The dog's NAME is LASSIE.
L+H* L- H* L-L%

3 The WordNet Lexical Database

WordNet is a large on-line English lexical database, based on theories of human lexical memory and comprised of four part-of-speech categories: nouns, verbs, adjectives, and adverbs (Miller et al., 1993). Within each category, lexical meaning is represented by synonym sets (synsets) organized around semantic relationships. Polysemous words are represented by multiple synsets, one for each word sense. The release used in this work, WordNet 1.5, contains a total of 91,591 synsets and 168,135 word senses (Miller, 1995).

Types of semantic relationships between synsets vary by category. The basic structure of each is discussed briefly below.

3.0.1 Nouns

The nouns category is the largest and semantically richest of the four. It contains 60,557 synsets, grouped into 25 different topical hierarchies. Synsets in each hierarchy are organized using hypernymy/hyponymy (IS-A) relationships. The noun hierarchies also include antonymy and three types of meronymy/holonymy relationships (PART-OF, MEMBER-OF, MADE-OF). Meronyms are typically defined at the level of basic concepts in the hierarchies.

3.0.2 Verbs

Verbs currently comprise 11,363 synsets in WordNet, divided into 15 categories based on semantic criteria. The primary semantic relationships for verbs in WordNet are lexical entailment (e.g. snoring ENTAILS sleeping) and hypernomy/hyponymy. Verb hierarchies also in-

[4] WordNet can capture some inferences, but is unable to account for a complex relationship like this one.

clude troponymy (MANNER-OF) relationships, and to a lesser extent, antonymy and causal relationships. Generally, verb hierarchies are much shallower with higher branching factors than noun hierarchies, but like nouns, verbs exhibit basic concept levels at which most troponyms are defined.

3.0.3 Adjectives

WordNet contains 16,428 synsets of adjectives divided into descriptive and relational types, and a small closed-class of reference-modifying adjectives. Descriptive adjectives are organized around antonymy, and relational adjectives according to the nouns to which they pertain. WordNet also encodes limitations on syntactic positions that specific adjectives can occupy.

3.0.4 Adverbs

Adverbs make up the smallest of the four categories, with a total of 3243 synsets. Adverbs are organized by antonymy and similarity relationships.

4 Implementation

An overview of the system architecture is shown in Figure 2. Following (Cahn, 1994; Cahn, 1997), text files are first parsed by the NPtool noun phrase parser, which identifies noun phrases and tags each word with morphological, syntactic, and part-of-speech information (Voutilainen, 1993). The preliminary processing module then adds gender information for proper names, resolves ambiguous tags, and reformats the text for further processing.[5] Next, the previous mention, contrast, and theme modules assign pitch accents, phrase accents, and boundary tones, using WordNet to identify sets of synonyms and contrastive words. Finally, the annotated text is re-formatted for the TrueTalk speech synthesizer (Entropic Research Laboratory, 1995). Additional implementation details for the accent assignment modules are provided below.

4.1 Givenness Identification

The first of the three accent assignment modules assigns pitch accents to words using the following given/new strategy:

For each word W,

1. If W is a noun, verb, adjective, or adverb, and W $\notin$ history( ), and W $\notin$ equiv(x), for any x $\in$ history( ):
   (a) tag W as a focused item
   (b) add W to history( )
   (c) create equiv(W)
2. If W is a noun, verb, adjective, or adverb, and W $\in$ equiv(x), tag W as inferable.[6]

The history and equivalence lists are reset at each paragraph boundary. Matches are limited to words belonging to the same part-of-speech category, relying only on word roots.

Equivalence (synonym) sets are created from semantic relationships for each WordNet category as follows:

1. Nouns: equiv(W) = union of hypernyms and synonyms for all synsets of W. The number of hypernym levels used for each sense is determined by searching for the existence of meronyms on the current level, climbing the hypernym tree until a level containing meronyms is found, or the root is reached. If no meronyms are found, then (1/4 × depth of W synset) levels are used.[7]
2. Verbs: equiv(W) = union of hypernyms, synonyms, and entailments for all synsets of W. Only one level of hypernyms is included.[8]
3. Adjectives and adverbs: equiv(W) = synonyms for all synsets of W.

[5] Gender resolution is performed via simple lookup using the CMU Artificial Intelligence Repository Name Corpus (Kantrowitz, 1994). Ambiguous parses are resolved using a set of heuristics derived from analysis of NPtool output.

[6] Items tagged as inferable by this step are realized by less prominent pitch accents than items tagged as focused, reflecting their status as not *explicitly* given.

[7] The present approach to identifying a "basic" concept level for nouns using meronymic relations is not the optimal solution. Many noun categories in WordNet do not include meronyms, and meronyms may exist at several levels within a hierarchy.

[8] Because verb hierarchies have a much higher branching factor, considering more than one level is generally impractical.

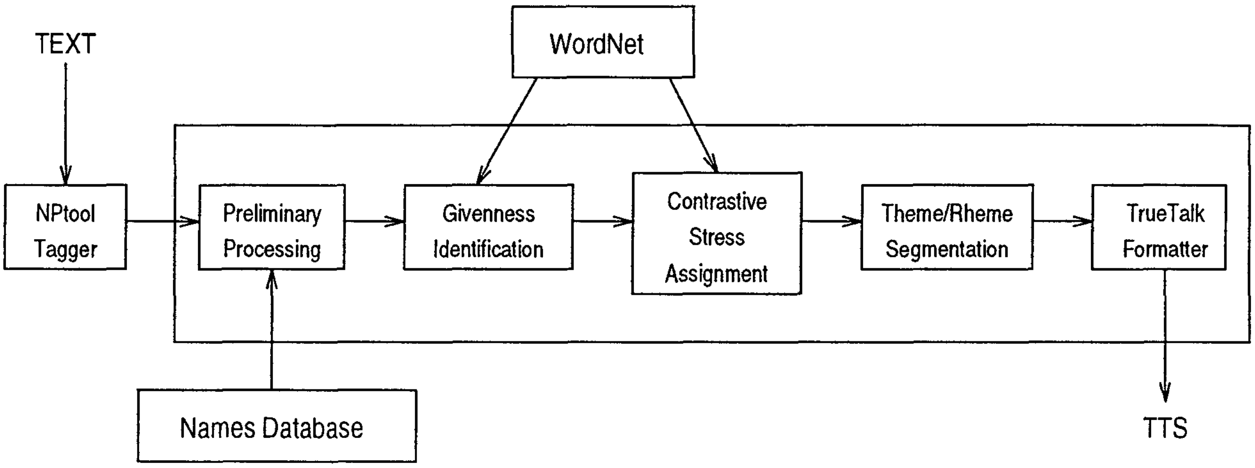

Figure 2: Architecture

Equivalence lists are ordered and searched from most common to least common sense of a word. The current implementation is limited to single word matches; no WordNet entries consisting of multi-word descriptions are included in the equivalence list.

4.2 Contrastive Stress Assignment

The second accent assignment module compares each open-class word (nouns, verbs, adjectives, and adverbs) with other words previously realized in the text to identify possible contrastive relationships. The top-level algorithm for assigning contrastive stress is shown in pseudocode in Figure 3.

Sets of possible contrastive words for nouns and verbs are determined by the hypernym/hyponym relationships in WordNet as follows:

1. Identify the set of immediate hypernyms, hyper(W), corresponding to each sense of W (synsets containing W).
2. For each h: h $\in$ hyper(W), identify the set of immediate hyponyms, hypo(h), such that W $\in$ hypo(h).
3. The set of possible contrastive words is the union of hyponyms for each sense of W.

The contrastive sets for adjectives and adverbs are simply the union of antonyms for all word senses, as hypernym/hyponym relationships are not used in WordNet for either class.

```
foreach word W1 {
  for each word W2 on the history list
  (from most to least recent) {
    for each A: A ∈ contrast(W2) {
      if W1 equals A then {
        tag W1 for contrastive stress;
        end search;
      }
    }
  }
  if no contrast is found {
    add W1 to the history list;
    generate & store {x : x ∈ contrast(W1)};
  }
}
```

Figure 3: Contrastive Stress Assignment

All contrastive sets generated are ordered and searched from the most common to least common sense of a word. The present implementation is limited to single word searches.

There are a number of shortcomings in the present implementation of contrastive stress assignment. The first is its failure to use textual information to facilitate identification of contrastive relationships. To rectify this situation, a search for keywords commonly used to indicate contrast (e.g. however, unlike, on-the-other-hand), as well as explicit negation (not) must be incorporated. Identifying parallel phrasing may also be useful in identifying contrastive relationships not encoded in WordNet (namely for non-antonymic contrasts between adjectives and adverbs).

4.3 Theme & Rheme Identification

The modules described above determine the second tier of the information structure—that is, which items are eligible for focus based on their new or contrastive status. The theme/rheme identification module is responsible for determining the primary information structure delineation of theme and rheme. Based on an automatic segmentation of utterances or parts of utterances into theme and rheme, we can apply the mapping of tunes described in Section 2 to decide which pitch accents to assign and where to place phrasal and boundary tones.

The automatic segmentation of utterances into theme and rheme is a difficult problem. Our preliminary approach is based on a number of heuristics, and generally performs quite well. Nonetheless, we expect this module to be substantially refined once we have concluded our empirical analysis of the Boston University radio news corpus (Ostendorf, Price, and Shattuck-Hufnagel, 1995).

The theme/rheme identification algorithm begins by trying to identify propositional constituents within utterances. As noted in Section 2, a single utterance may contain several clauses corresponding to several semantic propositions. Propositional constituents are centered around verb occurrences. The algorithm looks for verb complexes—contiguous stretches of text containing verbs, adverbs and some prepositions. Utterances are then divided into propositional constituents such that each contains a single verb complex. The algorithm also considers multi-word clauses that are set apart by punctuation, such as utterance-initial prepositional phrases, as separate propositional constituents.[9] This segmentation scheme is similar to Gussenhoven's division of utterances into focus domains (Gussenhoven, 1983).

Once propositional constituents have been determined, the algorithm applies a number of heuristics to sub-divide each into theme and rheme. We consider two possible segmentation points: before the verb-complex and after the verb-complex. The heuristics are as follows, where PreV, V and PostV correspond to the pre-verbal material, the verb-complex material and the post-verbal material respectively.

1. In the case where neither PreV, V nor PostV contains focused items:
   theme = [PreV]
   rheme = [V PostV]
   Accent V material.

2. If PreV and V contain focused items, but PostV does not:
   theme = [PreV]
   rheme = [V PostV]

3. If PreV and PostV contain focused items, but V does not:
   theme = [PreV V]
   rheme = [PostV]

4. If V and PostV contain focused items, but PreV does not:
   theme = [PreV V]
   rheme = [PostV]

5. If PreV, V and PostV all contain focused items:
   theme = [PreV V]
   rheme = [PostV]

6. If PreV contains focused items, but V and PostV do not:
   rheme = [PreV]
   theme = [V PostV]

7. If V contains focused items, but PreV and PostV do not:
   theme = [PreV]
   rheme = [V PostV]

[9] Note that we work with the part-of-speech output of NPtool rather than a complete parse tree. While this presents a number of difficulties for dividing utterances into propositional constituents, it allows us more freedom in sub-dividing those propositional constituents into theme and rheme. That is, our program can produce prosodic phrases, such as those shown in Figure 1, that are orthogonal to traditional syntactic structures.

8. If PostV contains focused items, but PreV and V do not:
   theme = [PreV V]
   rheme = [PostV]

Note that these heuristics encode a preference for thematic phrases to precede rhematic phrases, but do not always dictate such an ordering. Also, note that the heuristics allow thematic phrases to sometimes contain focused items. This is in accordance with our observation in Section 2 that themes need not contain only background material.

Based on the theme/rheme identification heuristics, we map **L+H*** accents onto focused items in themes and **H*** accents onto focused items in rhemes. **L-** phrasal tones are placed at theme and rheme boundaries. When theme or rheme phrases are also marked by punctuation, appropriate boundary tones and pauses are also inserted (e.g. **H%** for comma delimited phrases).

5 Results and Conclusions

The system was designed and debugged using a set of five single-paragraph texts. It was then tested using several new single-paragraph texts, excerpted from news articles and encyclopedia entries. Sample output is shown in Figures 4 and 5, where prominence, defined as a multiplier of the default nuclear accent, is shown directly below the associated pitch accent.

These preliminary test results indicate using information structure in conjunction with WordNet can produce intonational patterns with context-appropriate variation in pitch accent type and prominence. In general, **L+H*** accents occur on items deemed to be thematic, and **H*** accents occur on rhematic items. WordNet proved to be fairly successful at identifying words which were "given" via inference, thus allowing the program to correctly reduce the pitch accent prominence assigned to these words. For example, in Figure 4, the prominence of the pitch accent on "achievement" is lowered because of its relationship to "feat." In Figure 5, the prominence of the accent on "soil" is lowered because of its relationship to "ground." To a lesser extent, WordNet was also able to identify appropriate contrastive relationships, such as the relationship between "difficult" and "easy" in Figure 5. Consequently, our program places a slightly more prominent accent on "difficult" than it would have if "easy" had not occurred within the same segment.

While quite encouraging, these preliminary results have also identified many opportunities for improvement. The current implementation is limited by the absence of a full parse tree. It is also limited by the current heuristic approach to phrase segmentation, and therefore often produces **L-** phrasal tones in improper places. Substituting better tools for both parsing and phrase segmentation would improve the overall performance.

The system's accuracy level for WordNet synonym and contrast identification can be improved in two ways: by incorporating word sense disambiguation, and by using a more sophisticated approach for generating a "match." Presently, WordNet results are searched in order of most common to least common word senses, thus biasing matches towards common word senses, rather than determining the most likely context. Incorporating a sense disambiguation algorithm, such as that discussed in (Resnik, 1995), is a logical next step. Word matches are also limited to comparisons between individual words within a single part-of-speech category. Extending consideration to adjacent words and semantic roles would greatly reduce the number of spurious matches generated by the system.

Another area for improvement concerns the prominence of pitch accents. Based on our preliminary results, we believe that the **L+H*** accents should be somewhat lower than those shown in Figures 4 and 5. Once we have completed our analysis of the Boston University radio news corpus (Ostendorf, Price, and Shattuck-Hufnagel, 1995), we expect to modify the accent prominences based on our findings.

Our assessment of system performance is based on human listeners qualitative measurements of the "comprehensibility" of output from our system in comparison with the standard

```
The cloning of an adult sheep in Scotland seems likely to spark an
      L+H*         L+H*  L+H*      L+H*    L+H* L- H*       H*
       1.1         1.1   1.1       1.1     1.1     1.1      1.1
intense debate about the ethics of genetic engineering research in
  H*      H*               H*        H*        H*         H*
  1.1     1.1              1.1       1.1       1.1        1.1
humans.   But experts agree that, however the debate is resolved, the
  H* L-L%      L+H* L-  H*   L-H%    H*            L-   L+H* L-H%
  0.7          1.3      1.1          1.1                1.1
genie is irretrievably out of the bottle.  The unprecedented feat was
L+H*         L+H*      L+H* L-      H* L-L%         L+H*     L+H*
1.1          1.1       1.1          1.1             1.1      1.1
considered by many scientists to be impossible because of the
L+H*    L-     H*      H*              H*        H*
1.1            1.1     1.1             1.1       1.1
technical difficulties involved in nurturing genetic material and
    H*         H*         H*          H*       H*       H*  L-
    1.1        1.1        1.1         1.1      1.1      1.1
prompting it to grow into an intact organism.  Many more scientists
  L+H*  L-       H*           H*       H* L-L% L+H* L+H*
  1.1            1.1          1.1      1.1     1.1  1.1
have considered it an ethically dubious goal  because the achievement
                 L-      H*        H*    H* L-                L+H*
                         1.1       1.1   1.1                  0.7
theoretically opens the door to cloning humans, a possibility fraught
    L+H*      L+H* L-     H*            H* L-L%       H*        H*
    1.1       1.1         1.1           0.7           1.1       1.1
with moral ambiguities.
       H*      H*  L-L%
       1.1     1.1
```

Figure 4: Results for an excerpt from the Los Angeles Times, February 24, 1997

TrueTalk output. Although adequate for preliminary tests, better performance measurements are needed for future work. Possibilities include testing listener comprehension and recall of speech content, and comparing the system's output with that of several human speakers reading the same text.

Acknowledgments

The authors would like to thank Janet Cahn for many helpful suggestions, and particularly for her advice on selecting WordNet, NPtool and the Boston University radio news corpus. We would also like to thank Ken Haase for facilitating many helpful discussions, and the Maui High Performace Computing Center (MHPCC) for use of computing facilities. This research was funded by the generous sponsors of the MIT Media Laboratory.

References

Cahn, Janet. 1994. Context-sensitive prosody for text-to-speech synthesis. Technical Report 94-02, MIT Media Laboratory.

Cahn, Janet. 1997. *Prosody as a Consequence of the Capacity and Contents of Memory.* Ph.D. thesis, Massachusetts Institute of Technology. Forthcoming.

```
Termites are frequently classed as pests.   Although only  10  percent
 L+H*           L+H*     L+H*   L- H* L-L%           L+H* L+H*  L+H*
 1.1            1.1      1.1       1.1               1.1  1.1   1.1
of the known species have    destructive habits,   these species may do
        L+H*   L+H*  L+H* L-      H*      H*  L-L%  L+H*             L+H*
        1.1    1.1   1.1          1.1     1.1       1.1              1.1
great damage.  Subterranean termites,   which enter  wooden structures
  H*   H* L-L%       H*            L-H%       L+H* L-  H*      H*
  1.1  1.1           1.1                      1.1      1.1     1.1
through the ground,   as they need to maintain   contact with the soil's
              H* L-H%         L+H*       L+H* L-  H*               H*
              1.1             0.9        1.1      1.1              0.7
moisture,   are fairly  easy to control.   Insecticides can be placed in
 H*    L-H%     L+H* L-  H*      H*   L-L%    L+H*              L+H*  L-
 1.1            1.1      1.1     1.1          1.1               1.3
trenches dug around the structure to be protected.   Materials such as
  H*     H*                              H*    L-L%    L+H*    L+H*
  1.1    1.1                             1.1           0.9     1.1
pressure treated wood and reinforced concrete are impervious to
  L+H*    L+H* L- H* L-       L+H* L-  H*     L-    H*
  1.1     1.3     1.1         1.1      1.1          1.1
termites and make   safe foundations.  Dry wood termites,  however,  nest
       L-    L+H* L- H*      H*   L-L% L+H*           L-H% L+H* L-H% L+H%
             0.7     1.1     0.7       1.1                 1.1       1.1
within the wood they feed on and are much more difficult to control;
     L-              H*   L-         L+H*   L-  H*                L-L%
                     1.1             1.1        1.3
fumigation has proved to be the best technique.
    L+H*       L+H*      L-      H*      H*   L-L%
    1.1        1.3              0.7     1.1
```

Figure 5: Results for an excerpt from the Britannica online service

Entropic Research Laboratory, 1995. *TrueTalk Reference Manual.*

Gussenhoven, Carlos. 1983. *On the Grammar and Semantics of Sentence Accent.* Foris, Dodrecht.

Hirschberg, Julia. 1990. Accent and discourse context: Assigning pitch accent in synthetic speech. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 952–957.

Hirschberg, Julia. 1993. Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence*, 63:305–340.

Horne, M., M. Filipsson, M. Ljungqvist, and A. Lindstrom. 1993. Referent tracking in restricted texts using a lemmatized lexicon: Implications for generation of prosody. In *Proceedings of Eurospeech '93*, volume 3, pages 2011–2014, Berlin.

Kantrowitz, Mark. 1994. CMU artificial intelligence repository name corpus. http:// almond.srv.cs.cmu.edu/afs/cs.cmu.edu/project /ai-repository/ai/html/air.html.

Miller, G.A. 1995. WordNet version 1.5, Unix release notes. Technical report, Cognitive Science Laboratory, Princeton University.

Miller, G.A., R. Beckwith, C. Fellbaum, D. Gross, and C. Miller. 1993. Introduction to WordNet: an on-line lexical database, five papers on WordNet. Technical report, Cognitive Science Laboratory, Princeton University.

Monaghan, Alex. 1991. *Intonation in a Text-to-Speech Conversion System*. Ph.D. thesis, University of Edinburgh.

Monaghan, Alex. 1994. Intonation accent placement in a concept-to-dialogue system. In *Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*, pages 171–174, New Paltz, NY, September.

Ostendorf, M., P.J. Price, and S. Shattuck-Hufnagel. 1995. The Boston University radio news corpus. Technical Report ECS-95-001, Boston University.

Pierrehumbert, Janet and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip Cohen, Jerry Morgan, and Martha Pollock, editors, *Intentions in Communication*. MIT Press, Cambridge, MA, pages 271–312.

Pitrelli, John, Mary Beckman, and Julia Hirschberg. 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, Yokohama, September.

Prevost, Scott. 1995. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. Ph.D. thesis, University of Pennsylvania. IRCS Report 96-01.

Prevost, Scott. 1996. An information structural approach to monologue generation. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, pages 294–301, Santa Cruz.

Prevost, Scott and Mark Steedman. 1994. Specifying intonation from context for speech synthesis. *Speech Communication*, 15:139–153.

Prince, Ellen F. 1981. Towards a taxonomy of the given/new distinction. In P. Cole, editor, *Radical Pragmatics*. Academic Press, London, pages 223–255.

Resnik, Philip. 1995. Disambiguating noun groupings with respect to wordnet senses. In *Third Workshop on Very Large Corpora*, Cambridge, MA.

Steedman, Mark. 1991. Structure and intonation. *Language*, pages 260–296.

Terken, Jacques and Julia Hirschberg. 1994. Deaccentuation of words representing 'given' information: Effects of persistence of grammatical function and surface position. *Language and Speech*, 37(2):125–145.

Voutilainen, Atro. 1993. NPtool: a detector of English noun phrases. In *Proceedings of the Workshop on Very Large Corpora*.