



Discourse Structural Constraints on Accent in Narrative *

Christine H. Nakatani
Aiken Computation Laboratory, Harvard University
33 Oxford St., Cambridge, MA 02138, USA
Email: chn@das.harvard.edu

ABSTRACT

This paper examines the relationship between discourse structure and intonational prominence or pitch accent. It is argued, based on the distribution of pitch accent in spontaneous narrative speech, that accent function must be interpreted against a dynamic background of linguistic factors, including grammatical function, lexical form, and discourse structural constraints that affect the attentional status of referents. Modeling of the interactions among these factors should enable richer prosodic variation in speech generation systems.

INTRODUCTION

The hierarchical structure of discourse influences accent decisions in systematic ways. Previous studies have shown, for example, that the accentuation of referring expressions can be correlated with discourse structural properties (Terken 1984), and that taking discourse structure into account can improve the performance of pitch accent assignment algorithms (Hirschberg 1991).

The above findings await integration in a coherent model of pitch accent assignment that takes into account other linguistic factors that influence accentuation. This paper takes a step in this direction, by examining the interactions of discourse structural constraints with accentuation, grammatical function, and lexical form. Previous findings are expanded upon to account for the distributions of pitch accent in unrestricted spontaneous narrative speech. The new analyses of accent function are formally cast within the computational discourse modeling framework of Grosz and Sidner 1986 (hereafter G&S).

NARRATIVE STUDY

A total of 481 animate referring expressions in a 20-minute long American English unrestricted spontaneous narrative monologue were analyzed with respect to grammatical function, lexical form of expression and accentuation. Accented expressions in this study contain heads marked by either H* or a complex pitch accent in Pierrehumbert's system of English intonation (Pierrehumbert 1980). Table 1 gives the percent accented for the most frequently occurring lexical forms and grammatical functions. The narrative data obey the general tendency for pronouns

*This work was partially supported by a National Science Foundation (NSF) Graduate Research Fellowship and NSF Grant No. 447557525 at Harvard University. The author thanks Barbara Grosz and Julia Hirschberg for valuable discussion, and an anonymous reviewer for helpful comments.

PERCENT ACCENTED						
	Subject		Direct Object		TOTAL	
	%	N	%	N	%	N
3rd person pronouns	23%	25/111	7%	1/15	21%	26/126
Explicit lexical forms	89%	58/65	50%	14/28	77%	72/93
TOTAL	47%	83/176	35%	15/43	45%	98/219

Table 1: Percentage of accented referring expressions

to be unaccented, and content words, accented. Closer examination, however, reveals an asymmetric interaction of grammatical function and lexical form. Subjects are accented only 23% of the time as pronouns, but are regularly accented as explicit forms ($p < .00001$, $\chi^2 = 73.2$, $df = 1$). Direct objects, on the other hand, are unaccented half of the time as explicit forms, but are rarely accented as pronouns ($p < .005$, $\chi^2 = 8.1$, $df = 1$).

Clearly, grammatical function and lexical form alone cannot be used to reliably predict accentuation. The notion of discourse-old/discourse-new status is also of little value in predicting accent for these problematic classes, since all of the pronouns in the narrative, subject or otherwise, refer to discourse-old entities, while only one quarter of the accented direct object explicit forms refer to discourse-new entities.

MODELING OF ATTENTIONAL STATE

It is proposed that the distributions of pitch accent can be explained by the communicative functions that accent serves in various linguistic configurations, as summarized in Table 2. The proposed functions of accent share in common the fact that each serves to manipulate the dynamic record of the activated or salient entities in a discourse.

In the computational discourse modeling framework of G&S, the notion of salience is formalized by attentional focusing mechanisms that are claimed to underly discourse processing in general.¹ The present analyses require rudimentary definitions of four attentional statuses provided by the G&S model: primary local focus or Cb, secondary local focus, immediate global focus and non-immediate global focus.

At the local level of attentional focusing, each utterance contains one most centrally salient entity in primary local focus, called the Cb. Other entities realized by expressions in the utterance are in secondary local focus, and are candidates to become the Cb of the next utterance. The global level of attentional state is modeled as a LIFO stack of focus spaces. Each focus space dynamically represents the entities and relations that are salient within a single discourse segment. An empty focus space is pushed onto the stack when a segment begins; entities are recorded in the focus space as the discourse advances until the discourse segment closes and the focus space is popped from the focus stack. Those entities represented in the top focus space on the stack are in immediate global focus. Entities represented elsewhere on on

¹G&S propose three inter-related components of discourse structure: intentional structure, linguistic structure, and attentional state. Briefly, a discourse is comprised of coherent discourse segments whose hierarchical relationships are determined by intentional structure and realized by linguistic structure. Discourse processing proceeds at two levels, global and local. The global level concerns relationships among discourse segments, while the local level concerns relationships among utterances within a single segment.

	<i>Linguistic Factors</i>	<i>Discourse Function</i>
ACCENTED EXPRES- SIONS	SBJ pronoun SBJ explicit form DOBJ explicit form	Shift local attention to new <i>Cb</i> Introduce new global referent as <i>Cb</i> Introduce new global referent
UNACCENTED EXPRES- SIONS	DOBJ explicit form DOBJ pronoun SBJ pronoun	Maintain referent in global focus Maintain non- <i>Cb</i> referent in secondary local focus Maintain <i>Cb</i> referent in primary local focus

Table 2: Discourse functions of accent

the stack are in non-immediate global focus.

For the narrative study, the global and local focusing statuses of the referring expressions in the narrative were determined based on a discourse segmentation by the author (following procedures described in Grosz and Hirschberg 1992) and centering constructs (determined by hand using the segmentation analysis and the centering rules as defined in Grosz et al. 1983).

MANIPULATIONS OF GLOBAL FOCUS

Space permits only partial discussion of the accentuation behaviors summarized in Table 2, namely the problematic cases of direct object explicit forms. (For discussion of accented subject pronouns, see Nakatani 1993). Analysis of the direct object explicit forms shows that accentuation is generally determined by the global focus state. 80% of these expressions are cases of first mention since the last discourse segment boundary, although three quarters of these are also references to discourse-old entities. The accentuation of proper names in particular, which make up 83% of explicit forms in the narrative, has presented a puzzle for prosody researchers. The preponderance of initial mentions in a segment seems to accord with the idea that the newness of an entity relative to its discourse context can be communicated by the use of an explicit lexical form reintroducing the entity, including proper names (Sanford et al. 1988). Such reintroductions often bear pitch accent (Hirschberg 1991). On the other hand, it has been predicted that proper names should often be unaccented because their very use presumes familiarity with the named entity on the part of the listener (Ladd 1980, p. 91). These observations can be reconciled by observing that the choice of lexical form reflects

INTRODUCTION OF NEW GLOBAL REFERENT:

[Follows segment about Masson]

and he [Masson] approached PAUL EISSLER, head of the Freudian archives based in New York

MAINTENANCE OF IMMEDIATE GLOBAL REFERENT:

[Segment-initial utterance, follows sister segment about Peter Swales]

but you see ANNA FREUD didn't like Peter Swales [L*]

Figure 1: Examples of direct object explicit forms (caps on relevant NPs indicate accent)

the speaker's need to reintroduce referents when a new discourse segment begins, while the choice of accentuation conveys distinct shades of newness/givenness that reflect the recent global focusing history of the discourse. In the narrative, accent on direct object explicit forms marks the introduction of entities into a segment's focus space; crucially, these entities do not occur in the sister segment or immediately embedding segment. Lack of accent marks the reintroduction of entities recently in global focus immediately following a focus space pop or push. The reintroduced entity crucially must be or have just been in immediate global focus. Examples of accented and unaccented direct object explicit forms appear in Figure 1.

This analysis of direct object explicit forms extends a previous hypothesis that reference resolution proceeds differently for accented and unaccented expressions, namely that listeners assume that an unaccented expression refers to a member of a "restricted set of activated entities" in the discourse context, while the interpretation of an accented expression is not constrained in this manner (Terken and Nootboom 1987, p. 148). The notion of activated entities is formally applied in the narrative study in terms of the structured contents of the global focus stack.

CONCLUSION

The narrative study attempts to interrelate previous findings about factors influencing accentuation, working towards a unified theory of the various communicative roles of accent within a discourse processing framework. The stated hypotheses concerning accent function will be tested on direction-giving monologues, which contain a fuller variety of lexical forms of referring expressions. Distributions of pitch accent type as well as quantitative measures of relative intonational prominence are areas for future investigation on this new corpus. Finally, the principles of pitch accent assignment resulting from these studies will be implemented in a message-to-speech system, in which discourse structure and meaning can be more directly encoded to generate speech of moderate prosodic and discourse structural complexity.

REFERENCES

- Grosz, B. J., Joshi A. K. and S. Weinstein 1983. "Providing a unified account of definite noun phrases in discourse", *Proceedings of 21st Meeting of the Association for Computational Linguistics*, Cambridge, MA.
- Grosz, B. J. and J. Hirschberg 1992. "Some intonational characteristics of discourse structure", *Proceedings of the International Conference on Spoken Language Processing*, Banff, pp. 429-432.
- Grosz, B. J. and C. Sidner 1986. "Attention, intentions, and the structure of discourse", *Computational Linguistics*, 12(3).
- Hirschberg, J. 1991. "Pitch accent in context: Predicting intonational prominence from text", *Artificial Intelligence*, 63(1/2).
- Ladd, D. R. 1980. *The Structure of Intonational Meaning*, Indiana University Press, Bloomington.
- Nakatani, C. 1993. "Accenting on pronouns and proper names in spontaneous narrative. *Proceedings of the European Speech Communication Association Workshop on Prosody*, Lund, Sweden.
- Pierrehumbert, J. 1980. "The phonology and phonetics of English intonation", Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge.
- Sanford, A., Moar, K. and S. Garrod 1988. "Proper names as controllers of discourse focus", *Language and Speech*, Vol. 31, pp. 43-56.
- Terken, J. 1984. "The distribution of pitch accents in instructions as a function of discourse structure", *Language and Speech*, Vol. 27, Part 3, pp. 53-73.
- Terken, J. and S. Nootboom 1987. "Opposite effects of accentuation and deaccentuation on verification latencies for Given and New information", *Lang. and Cognitive Processes*, 2(3/4).