Effect of Genre, Speaker, and Word Class on the Realization of Given and New Information

Agustín Gravano, Julia Hirschberg

Department of Computer Science
Columbia University, New York City, NY, USA
{agus,julia}@cs.columbia.edu

10.21437/Interspeech.2006-208

Abstract

There is much evidence in the literature that speakers tend to deaccent discourse-given entities, while accenting new ones. However, speakers do not always follow this simple strategy and the causes for such variation are not yet well understood. In this paper, we describe several new forms of variability in the relationship between given/new information and accenting behavior, variation due to individual differences and to word class. We present results indicating that different speakers have different strategies for making new words prominent. We analyze two word-classes – nouns and verbs – in a corpus of spontaneous and read direction-giving monologues, and show that speakers use different combinations of pitch, intensity and inter-word pauses to distinguish between given and new information. Most interestingly, we find that in both genres all speakers tend to produce given verbs with higher intensity than new verbs.

Index Terms: prosody, information status, given/new information, accenting.

1. Introduction

There is considerable evidence in the literature that speakers of American English tend to ***deaccent*** discourse-given entities, while ***accenting*** new ones [1][2][3][4]. Usually, the information status of discourse entities is described using models such as [2][3][5], which are difficult for labelers to label reliably and still more difficult to implement automatically. However, simpler models have also been shown to effectively model the given/new distinction, for applications like speech synthesis [6].

In this paper, we employ a simple definition of information status, and look for correlations between new/given words and (semi-)automatically extractable acoustic features. We want to examine in particular how different speakers produce given and new information and to see whether their productions are influenced by word-class.

2. The Boston Directions Corpus

The current investigation makes use of a corpus of spontaneous and read speech, the Boston Directions Corpus [7]. This corpus comprises elicited monologues produced by four non-professional speakers, three male (S1, S2 and S4) and one female (S3), who were given written instructions to perform a series of nine increasingly complex direction-giving tasks. Speakers first explained simple routes such as getting from one station to another on the subway, and progressed gradually to the most complex task of planning a round-trip journey from Harvard Square to several Boston tourist sights. The speakers were provided with various maps, and could write notes to themselves as well as trace routes on the maps. For the duration of the experiment, the speakers were in face-to-face contact with a silent partner (a confederate) who traced on her map the routes described by the speakers. The speech was subsequently orthographically transcribed, with false starts and other speech errors repaired or omitted; subjects returned several weeks after their first recording to read aloud from transcriptions of their own directions. A total of 50 minutes of read speech and 66.6 minutes of spontaneous was collected, with speakers ranging from 7.9 to 17.9 minutes for the read tasks and 11.2 to 22.8 for spontaneous productions.

3. Methods

3.1. Information status

Instead of labeling our data using an information status model such as [3] or [4], we use a shallow definition of givenness. Following [6], we say that a word *w* is *given* if in the given task there is at least one previous occurrence of a word with the same stem; otherwise, we say that *w* is *new*.

3.2. Part-of-speech categories

We tagged all words for part of speech using the Brill Tagger [8] and collapsed the tags into the following classes (for a description of the tags, see [9]):

Noun = {NN,NNS,NNP,NNPS}
Verb = {VB, VBD, VBG, VBN, VBP, VBZ}
Adjective = {JJ, JJR, JJS}
Adverb = {RB, RBR, RBS, WRB}
Other = all other tags

Tables 1 and 2 show the distribution of new and given words in the corpus, by speaker and POS category.

	Speaker 1	Speaker 2	Speaker 3	Speaker 4
	new / given	new / given	new / given	new / given
Adj	36 / 98	56 / 138	14 / 43	42 / 117
Adv	70 / 84	185 / 156	17 / 51	128 / 125
Noun	351 / 310	572 / 594	232 / 290	383 / 484
Verb	139 / 178	241 / 500	64 / 142	176 / 295
Other	23 / 935	120 / 1565	28 / 575	74 / 1184

Table 1. *Distribution of new and given words per speaker and POS category (read data)*

	Speaker 1	Speaker 2	Speaker 3	Speaker 4
	new / given	new / given	new / given	new / given
Adj	35 / 94	61 / 142	14 / 40	41 / 106
Adv	68 / 85	202 / 158	17 / 56	134 / 127
Noun	379 / 314	598 / 622	242 / 308	419 / 529
Verb	139 / 185	266 / 509	68 / 151	209 / 295
Other	62 / 976	129 / 1665	66 / 601	174 / 1330

Table 2. *Distribution of new and given words per speaker and POS category (spontaneous data)*

These tables show a higher ratio of new to given nouns for Speakers 1 and 2 than for Speakers 3 and 4, although only Speaker 1 produces more new nouns than given. Ratios of new to given verbs are also highest for Speaker 1.

3.3. Acoustic Features

We used Praat [10] to extract pitch and intensity values for the corpus, in both cases with a sampling rate of 200 Hz. Inter-word pauses were extracted automatically from the orthographic alignment. We examined the following features for each word:

{Max, Mean, Min}Pitch: word maximum, mean, or minimum pitch.

{Max, Mean, Min}Pitch / TimeContext: ratio of the word maximum, mean or minimum pitch, and the mean pitch of its *time context*, defined as the word itself plus 1 sec preceding it and 1 sec following it.

{Max, Mean, Min}Pitch / WordsContext: ratio of the word maximum, mean or minimum pitch, and the mean pitch of its *words context*, defined as the word itself plus its 5 preceding and 5 following words (up to a limit of 5 seconds to each side).

{Max, Mean, Min}Pitch / IP: ratio of the word maximum, mean or minimum pitch, and the mean pitch of its intermediate phrase.[1]

PauseBefore: length of the silence before the word.

PauseAfter: length of the silence after the word.

4. Analysis and Results

In this section, we show how different speakers produced given and new items for different parts of speech, in terms of our acoustic features. We present results only for our Noun and Verb categories, since these were the only categories in which we found significant differences between productions.

Each cell in these tables represents the comparison of an acoustic variable (e.g., maximum pitch) between two sets of words: the given and the new words of a particular speaker (S1-S4). If a cell contains an "n", this means that the mean of the corresponding variable for the new words is significantly larger than for the given words. Conversely, if the cell contains a "g", then the given words have a larger mean for that variable than the new words. Finally, if the cell is empty, then no significant difference was found. In all cases, we performed a two-sided *t*-test on the means, and considered a result to be statistically significant when $p < 0.05$.

[1] Intermediate phrases are defined in the ToBI labeling conventions [11]. Automatic detection of IP boundaries is very difficult for human labelers [12] and even more so for automatic methods [13]. However, these variables are of interest from a linguistic perspective.

4.1. Nouns

In this section we compare the production of nouns across speakers, for both read and spontaneous speech.

	READ				SPON			
	S1	S2	S3	S4	S1	S2	S3	S4
MaxPitch								g
MeanPitch	n				n			g
MinPitch	n							g
MaxPch / TimeContext	n	n			n	n		
MeanPch / TimeContext	n	n	n		n			
MinPch / TimeContext	n				n			
MaxPch / WordsContext		n						
MeanPch / WordsContext	n	n			n			
MinPch / WordsContext	n							
MaxPitch / IP	n	n	n		n			
MeanPitch / IP	n	n	n		n		n	
MinPitch / IP	n				n			

Table 3. *Nouns, pitch, read and spontaneous data.*

	READ				SPON			
	S1	S2	S3	S4	S1	S2	S3	S4
MaxIntensity	n				n	g		g
MeanIntensity	n				n	g		g
MinIntensity						g	g	g
MaxInt / TimeContext	n		n		n		n	
MeanInt / TimeContext	n		n		n	g		g
MinInt / TimeContext								g
MaxInt / WordsContext			n		n		n	
MeanInt / WordsContext					n			g
MinInt / WordsContext							g	g
MaxIntensity / IP	n		n		n		n	
MeanIntensity / IP	n				n	g		
MinIntensity / IP	n							

Table 4. *Nouns, intensity, read and spontaneous data.*

	READ				SPON			
	S1	S2	S3	S4	S1	S2	S3	S4
PauseBefore		n	n		n	n	n	
PauseAfter		g			g	g		

Table 5. *Nouns, pause, read and spontaneous data.*

4.1.1. Read data

The left halves of Tables 3-5 indicate that speakers S1, S2 and S3 vary different combinations of pitch, intensity and pause when eliciting new nouns in read speech, all supporting the hypothesis that new entities are more prominent than old ones. No significant variation was observed for S4.

4.1.2. Spontaneous data

The right halves of Tables 3-5 show quite different patterns for spontaneous speech for three of our speakers. S1 exhibits a very similar pattern as for his read productions. There is almost no significant difference in S2's and S3's pitch, but now S2 clearly produces new nouns with lower intensity, while S3 produces them using an expanded intensity range. Surprisingly, S4's data suggests that he gives more prominence not to new nouns, but to given ones. This finding of significant differences in pitch prominence for ***given*** items is, to our knowledge, quite unusual and has not previously been reported in the literature. [2]

4.2. Verbs

Tables 6-8 show how each speaker produced given and new verbs, from read and then from spontaneous speech.

	READ				SPON			
	S1	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
MaxPitch							n	
MeanPitch	g	g					n	
MinPitch	g						n	
MaxPch / TimeContext	g						n	
MeanPch / TimeContext	g							
MinPch / TimeContext	g		n					
MaxPch / WordsContext							n	
MeanPch / WordsContext	g							
MinPch / WordsContext	g							
MaxPitch / IP								
MeanPitch / IP								
MinPitch / IP			n					

Table 6. *Verbs, pitch, read and spontaneous data.*

	READ				SPON			
	S1	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
MaxIntensity	g	g	g		g	g	g	g
MeanIntensity	g	g	g	g		g	g	g
MinIntensity				g			g	g
MaxInt / TimeContext	g	g	g		g	g		
MeanInt / TimeContext	g	g	g	g	g	g	g	g
MinInt / TimeContext				g				g
MaxInt / WordsContext	g	g	g		g	g	g	
MeanInt / WordsContext	g	g	g	g		g	g	g
MinInt / WordsContext				g			g	g
MaxIntensity / IP	g	g	g		g	g		
MeanIntensity / IP	g	g		g		g	g	g
MinIntensity / IP				g				g

Table 7. *Verbs, intensity, read and spontaneous data.*

[2] S4 uses a higher pitch for given nouns with respect to his overall pitch, but not to the word context. This means that he uses a higher pitch for the whole phrase and not just for the word, which could be due to reasons other than information status, such as discourse structure.

	READ				SPON			
	S1	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
PauseBefore		g		g				
PauseAfter			g					

Table 8. *Verbs, pause, read and spontaneous data.*

4.2.1. Read data

For read speech, the left half of Table 7 shows a marked uniformity in the tendency of all speakers to use ***greater*** intensity over given verbs than over new ones, contra their performance on nouns. However, while S1 uses higher pitch for given verbs, there is some use of higher pitch for new verbs as well. Also, while S2 and S4 produce a longer pause before given verbs, S3 produces a longer pause after given verbs.

4.2.2. Spontaneous data

Once again, we see that speakers uniformly produce given verbs with greater intensity than they do with new verbs (right half of Table 7).

5. Discussion

Overall, we observe that, as expected, new nouns are generally produced with higher pitch. However, surprisingly, all speakers, both in read and in spontaneous speech, produce **given verbs** with a **greater intensity** than new verbs. This seems to suggest that the given/new non-prominent/prominent default for verbs is reversed with respect to intensity. There is very little evidence though of a difference for pitch in verbs. Another observation across all speakers and both genres is that normally there are **longer pauses before new nouns**. There is also some evidence of **longer pauses after given nouns**.

In all cases, when comparing read to spontaneous speech, each of our four speakers shows the same direction of variation. That is, no speaker uses a higher value of a feature for a particular class in one genre, and a lower value in the other genre. However, there are major differences between speakers. S3 and S4 show more variation in spontaneous than in read speech for nouns and verbs, which could be explained by these speakers compressing their pitch and intensity ranges when reading. This is not true for S1 and S2. For some verbs in read speech, S1 uses a higher pitch and no variation in pause, while the other three speakers take an opposite strategy: with a longer pause either before or after the verb, and practically no variation in pitch. In both genres and for both nouns and verbs, there is almost no evidence of S4 varying his pitch due to word information status. At least in the case of verbs and spontaneous nouns, he does vary other features, however. Finally, S3 is the only speaker to increase her intensity range: for new nouns in spontaneous data.

5.1. Variation of intensity in verbs

Based only on these results, it would be unreasonable to attempt an explanation of the increased intensity observed in given verbs; further investigation is indeed required. However, looking at a few examples from the corpus might shed some light on this phenomenon.

In all of the following excerpts, first mentions of the highlighted verbs are uttered with a lower intensity than

subsequent mentions, both in the read and the spontaneous versions. On the other hand, their pitch level does not vary uniformly. The pound symbol (#) represents short pauses, typically shorter than one second.

(1) *you get out of the T stop you **cross** Massachusetts Avenue # when you get out of the T stop* [...] [74 more words describing the surroundings] *you wanna **cross** Mass Ave opposite that # there's usually a bunch of cabs and people standing around there # so then once you've **crossed** it you're on Harvard Yard*

(2) *so go in the building # **follow** the Infinite Corridor for basically as far as you can # one trick about this is that it's not quite straight so it runs for about six buildings and then takes a left turn and then a right turn and then goes a little farther # but keep **following** it* [...]

(3) *then you're right at the entrance to what is **called** the Infinite Corridor # and it's **called** the Infinite Corridor because it's this really long place* [...]

(4) *so you're going to have to **transfer** # you **transfer** by going to Government Center which is inbound* [...]

In (1), (2) and (3) the direct objects of verbs *'cross'*, *'follow'* and *'call'* are either deaccented or pronominalized in the second and third mentions. With the lack of other salient accented items in their respective phrases, it may not be surprising that the given mentions of these verbs are more prominent. And in (4), the increased prominence of the second mention of *'transfer'* might be due to its second mention in a different verb form from the infinitival form in which it was first mentioned, similar to findings of [4] that given nouns tend to be accented if they represent a different grammatical function from the first mention. Other research under more controlled conditions will be necessary to test these hypotheses generally.

6. Conclusions

The results presented in this paper represent preliminary findings on how speakers present discourse-given and discourse-new words, both in spontaneous and read speech. Using a shallow definition of information status, we classified all the words in the corpus according to its part of speech category, and found significant differences in several acoustic features between the sets of new and given words. While our findings support previous observations that speakers normally make new nouns more prominent than given ones, our results show that each speaker tends to use their own combination of acoustic features to signal such prominence. However, for verbs our findings reveal a different picture: for verbs in our corpus, the given/new non-prominent/prominent default seems to be reversed with respect to intensity, while there is very little evidence of difference for pitch. So, the relationship between intonational variation and information status appears to be heavily conditioned on part-of-speech. We are continuing to explore the role of syntactic and contextual features on the realization of given and new information on larger corpora.

7. Acknowledgements

This research was funded in part by NSF IIS-0307905. We thank Stefan Benus and the anonymous reviewers for helpful comments and suggestions.

8. References

[1] Chafe, W. L., "Language and consciousness", *Language*, Vol. 50, 111-133, 1974.

[2] Prince, E. F., "Toward a taxonomy of given-new information", *Radical Pragmatics*, Peter Cole (ed.), 223-255, The Academic Press, New York, 1981.

[3] Prince, E. F., "The ZPG letter: Subjects, definiteness, and information-status", in Thompson, S. and Mann, W. (eds.), *Discourse description: diverse analyses of a fund raising text*, John Benjamins B. V., Philadelphia/Amsterdam, 295-325, 1992.

[4] Terken, J. and Hirschberg, J., "Deaccentuation of words representing 'given' information: Effects of persistence of grammatical function and surface position", *Language and Speech*, Vol. 37 (2), 125-145, 1994.

[5] Gundel, J. K., Hedberg, N. and Zacharski, R., "Cognitive Status and the Form of Referring Expressions in Discourse", *Language*, Vol. 69 (2), 274-307, 1993.

[6] Hirschberg, J., "Pitch accent in context: predicting intonational prominence from text", *Artificial Intelligence*, Vol. 63, 305-340, 1993.

[7] Hirschberg, J. and Nakatani, C., "A prosodic analysis of discourse segments in direction-giving monologues", *ACL*, Santa Cruz, California, 286-293, 1996.

[8] Brill, E., "A simple rule-based part of speech tagger", *DARPA Workshop on Speech and Natural Language*, Morgan Kaufmann Publishers, Inc., San Francisco, California, 112-116, 1992.

[9] Marcus, M., Santorini, B. and Marcinkiewicz, M. A., "Building a large annotated corpus of English: The Penn Treebank", *Computational Linguistics*, Vol. 19 (2), 313-330, 1994.

[10] Boersma, P. and Weenink, D., "Praat: doing Phonetics by Computer", www.praat.org, 2001.

[11] Beckman, M. E. and Hirschberg, J., "The ToBI annotation conventions", Columbus, OH, Ohio State University, 1994.

[12] Pitrelli, J. F., Beckman, M. E. and Hirschberg, J., "Evaluation of prosodic transcription labeling reliability in the ToBI framework", *ICSLP*, Yokohama, Japan, Vol. 1, 123-126, 1994.

[13] Wightman, C. W. and Ostendorf, M., "Automatic labeling of prosodic patterns", IEEE Transactions on Speech and Audio Processing, Vol. 2 (4), 469-481, 1994.