

FACTORS AFFECTING THE GIVEN-NEW DISTINCTION IN SPEECH

Sarah Hawkins and Paul Warren

Department of Linguistics, Cambridge University, UK

ABSTRACT

Much attention has been paid to variation in acoustic properties depending on whether a word is "new" or "given" in a discourse. The hypothesis of this paper was that the given-new distinction is relatively unimportant in the perception of normal conversational speech. Selected words and CV fragments from those words were excised from conversations with 3 people and their intelligibility was measured. Sentence stress, the particular consonant involved and individual speaker characteristics all affect intelligibility more than the given-new distinction.

1. INTRODUCTION

Experiments show that the information a word contributes to a discourse can affect its intelligibility: more predictable words tend to be spoken less clearly than less predictable words. Predictability that has been shown to affect intelligibility includes the meaning and grammar [6], and whether the word has been used before in the discourse [2] - the so-called new-old, or given-new, distinction.

These differences in intelligibility are statistical tendencies: not all words are affected, and some of the differences are small. Moreover, whereas some studies find differences in acoustic measurements that correlate with intelligibility differences, others find no differences in the same parameters, albeit in different languages [cf. 2; 4; 5].

If the given-new distinction has a significant influence on the intelligibility of all speech, there would be important consequences for models of both human

and machine speech recognition. However, this paper reports preliminary work intended to investigate the possibility that "given versus new" is too simple a distinction to be useful for normal conversational speech.

One challenge in studying the given-new distinction is defining what is "old" information. Most studies treat the first instance of a word as new, and later instances as old, or given. While this may be appropriate in an analysis of the discourse, it is unlikely to be appropriate for predicting the intelligibility of individual words or parts of words in ordinary conversations. A second or later word may be spoken in isolation, or with contrastive stress, for example, both of which might be expected to increase rather than reduce its intelligibility. We do not question that predictability is one factor that can affect intelligibility. But we do suggest that in normal conversational speech, the given-new distinction has only a small effect on intelligibility; other factors will be at least as influential.

Patterns of intelligibility are likely to depend on the types of discourse and speech being analysed. Large intelligibility effects due to the given-new distinction have tended to be found with speech that has been controlled for several aspects of linguistic context, or with tasks where clarity of speech and style of presentation are crucial [1, 2]; even here, intelligibility also varies with the information content of the repeated word and the experience of the speaker [1].

Fluent reading of texts may give a distorted view of the prevalence of given-

new distinctions in speech. Texts designed to elicit such differences in intelligibility are likely to produce them. But these differences may be much less likely to occur in normal conversational speech, which typically has shorter and less grammatically complex phrases. Hunnicutt's [4] finding that a greater intelligibility effect arises with long sentences typical of the written but not the spoken language supports this view.

Word intelligibility is also likely to be influenced by phonetic factors. The prosodic context has already been mentioned. Differences due to segmental-phonetic structure could depend on the acoustic properties of the sounds involved and/or to the phonological inventory of the particular language. For example, stridency is normally a robust acoustic property, and the range of possible articulations for a strident sound is fairly small. Thus stridency involves *relatively* little spectral variation even in casual speech. For languages in which a strident-nonstrident distinction is phonemically contrastive, then, strident sounds might be expected to retain a high level of intelligibility in most contexts.

A phonetic difference that is mainly dependent on phonological space is the leniting of velar stops in English. The only velar consonants in English are oral and nasal stops; so, since /ŋ/ can only be syllable-final, and the acoustic correlates of nasalization are fairly distinctive and distributed over time, leniting /g/ and /k/ is unlikely to pose problems for the listener. In contrast, alveolar stops share a crowded section of English phonological space, and typically are not unlike strident fricatives in some of their spectral properties. In comparable phonetic environments, then, we would expect velar stops to vary more than alveolar stops in manner of articulation.

2. EXPERIMENT

To examine the worth of these arguments, we collected from natural conversational speech repeated tokens of the same words spoken by different people. We then measured the intelligibility of the whole words and their medial consonant. The words were all bisyllabic and stressed on the first syllable. The medial consonant was (a) the sound of interest (b) where the

word became lexically unique, and (c) one of /d g ə s ʃ/.

Medial consonants were chosen so that, as far as possible, the immediate phonetic context was controlled for coarticulation effects. Medials also allow the possibility of presenting CV, VC, and VCV portions of the words to listeners for identification. Requiring the medial consonant to represent the word's uniqueness point greatly constrained the choice of words, but had the advantage that word identification would take place under similar conditions of lexical access [cf 7].

The choice of sounds was governed by the existence of suitable words and by the following considerations. 1. /s ʃ/ are strident; the others are not. 2. /g/ will vary in manner of articulation more than the others, so under comparable conditions its intelligibility should vary most. 3. The experimental manipulations and acoustic analyses are more straightforward for voiced than for voiceless stops [3]. 4. The fricative /θ/ resembles /s ʃ/ in that it is long (so could have an intelligibility advantage when excised from running speech), but it is nonstrident.

3. HYPOTHESES

Over the whole corpus:

1. First tokens of words and of medial consonantal fragments will not differ in intelligibility from second tokens. This will also be true for the subset of first and later tokens bearing nuclear stress.
2. Tokens with nuclear stress will be more intelligible than with secondary or no stress, regardless of how many times the word has been used in the conversation.

Isolated sounds will differ in intelligibility such that:

3. Strident (/s ʃ/) sounds will be more intelligible than other sounds overall, and later instances will be as intelligible as the first instance.
4. Because we expect /g/ to vary more than /d/, /g/ will be more likely to show variation due to the given-new contrast and to differences in sentence stress.
5. People will differ in the overall intelligibility of their speech and in how much it conforms to these predictions.

4. METHOD

The selected materials were sorted into four 'topics'. Two women and one man, speakers of Southern British English, each discussed them with the experimenters in a sound-treated room. The speakers all knew the experimenters, and spoke in relaxed conversational style. Pictures were used to stimulate and guide discussion towards the words we were looking for. In the vast majority of cases the experimental subjects were the first users of the words of interest.

The repeated experimental words selected from within each speaker's discussion of the relevant topic were: 1. the first production of the word; 2. the second production; 3. where possible, the next production contrasting in stress with the second token. In this paper, the third tokens are only used in comparisons of nuclear with other stress levels. The resulting 21 word sets were digitally excised from their fluent contexts and recorded onto digital audio tape for presentation to listeners.

For word identification, tokens were heard in white noise at a signal-to-noise ratio of 5dB above the average intensity of the speech (excluding silence). Each subject heard only one token of each test word, counterbalanced across nine versions (3 speakers x 3 repetitions). The ISI was 4s, during which subjects wrote down the word they had just heard. Each test list had between 17 and 19 words and was preceded by 6 practice words.

In a second task, fragments containing consonantal information were excised: for stops, the burst and following 80 ms; for fricatives, the frication period plus 40 ms of the following periodicity. No noise was added. Each listener heard all excised segments in one of two randomisations, preceded by a 6-item practice list. The ISI was 2s, with a longer ISI after every tenth item. Listeners wrote down the consonant(s) they heard.

90 students completed the word identification task (10 on each version); 10 further students took part in the consonant task. Both tasks were open response. Listeners heard the materials over headphones in a sound-treated room.

5. RESULTS

The predictions were tested using ANOVAs, with designs differing according to comparison. We summarise some of the more interesting results so far. Differences reported as significant achieved a probability of 0.05 or better.

Words. Following [2,3] a response was scored as correct only if the whole word was identified correctly. Our argument that conversational speech should show no *general* tendency for the new-given distinction to appear is supported by the finding of no overall effect for this factor in the intelligibility scores. In contrast to this, we find a clear effect of stress type: words carrying nuclear stress are significantly clearer than others (68% vs 50%). Taken together with the distribution of stress types in our sample, this goes a long way towards accounting for the lack of a new-given distinction. The *new* items almost all have nuclear stress (92%), and so do a large minority of the *given* (44%). Unsurprisingly, amongst the words carrying nuclear stress, there is no effect of new vs given. There were also no overall speaker differences for word intelligibility.

In an attempt to control for some of the variability in parameters other than that of new vs given, we chose a subset of materials with comparable phonetic make-up (one word, produced by all speakers, from each of the five sound types). In this subset *new* items are significantly more intelligible (78% vs 45%). However, it is possible that there is a confounding effect here of prosodic context, since 13 of 15 *new* items are in nuclear position, but 4 of 15 *given*. Further work is needed here.

Consonants. In scoring of the identification of consonants, we are interested primarily in place and manner: errors in voicing only are therefore counted as correct. As expected, we found significant effects of sounds and speakers. Strident fricatives achieved by far the best scores (*f*/: 91%, *s*/: 87%), with *d*/ and *g*/ intermediate (56% and 55%) and *θ*/ worst (19%). The stress effect found for the word task is replicated here, with significantly fewer errors for consonants from words bearing nuclear stress (66% vs 53%).

The *d*/ and *g*/ groups were evaluated further to compare the contrast in "stridency" and "phonological space" discussed above. The figure shows the predicted significant interaction of sound (*/d-g/*) with given-new, as well as main effects of speaker and given-new. On the whole, */g*/ loses intelligibility on repetition whereas */d*/ does not, but the effects are much greater for some speakers.

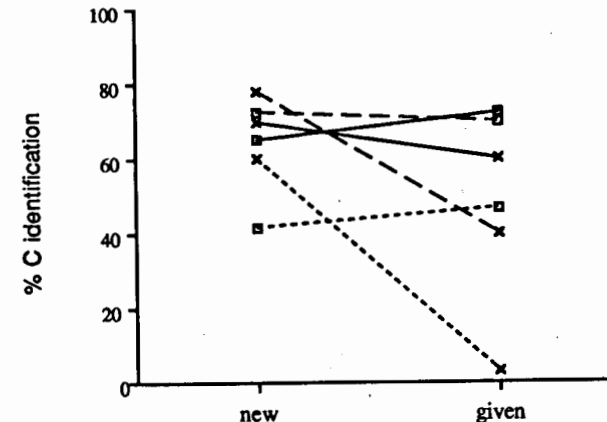


Figure: Consonant intelligibility for */d-/g/* substrs. Different line styles denote the 3 speakers. Crosses show */g/* and squares */d/* identification scores.

6. CONCLUSION

Our hypotheses regarding whole words (1 & 2) were supported by the general finding that sentence stress affects intelligibility more than the simple given-new distinction. The hypotheses for consonants were partially supported in that strident fricatives were always highly intelligible (3), and in that given-new differences appeared for */g/* but not */d/* (4). However, the sentence stress effect found for whole words did not appear for isolated consonants. Whereas speakers' whole words did not differ in intelligibility, there were large differences in the intelligibility of their isolated consonants (5). This finding suggests that individuals vary in how much they distribute acoustic cues within words; listeners' perceptual strategies must show the required flexibility [cf. 3,7].

7. REFERENCES

- [1] BARD, E.G., LOWE, A.J. & ALTMANN, G.T.M. (1989), "The effect of repetition on words in recorded dictations", *Eurospeech: Proc. 2nd Eur. Conf. on Speech Comm.*, 2, 573-576.
- [2] FOWLER, C.A. & HOUSUM, J. (1987), "Talkers' signalling of 'new' and 'old' words in speech and listeners' perception and use of the distinction", *J. Mem & Lang.*, 26, 489-504.
- [3] HAWKINS, S. (1989), "Reconciling trading relations and acoustic invariance", *Eurospeech*, 2, 682-685.
- [4] HUNNICUTT, S. (1987), "Acoustic correlates of redundancy and intelligibility", *Stockholm: KTH. STL-QPSR*, 2-3.
- [5] KOOPMANS-VAN BEINUM, F.J. & VAN BERGEM, D.R. (1989) "The role of 'given' and 'new' in the production and perception of vowel contrasts in read text and in spontaneous speech", *Eurospeech*, 2, 113-116.
- [6] LIEBERMAN, P. (1963), "Some effects of semantic and grammatical context on the production and perception of speech", *Lang. & Speech*, 6, 172-187.
- [7] WARREN, P. & MARSLER-WILSON, W.D. (1987), "Continuous uptake of acoustic cues in spoken word recognition", *Perception & Psychophysics*, 41, 262-275.