THE CONTROL OF INTELLIGIBILITY IN RUNNING SPEECH

E. G. Bard[1], C. Sotillo[1], A. H. Anderson[2], G. Doherty-Sneddon[2], and A. Newlands[2]

Human Communication Research Centre,
[1]Dept. of Linguistics, Edinburgh University, [2]Dept. of Psychology, Glasgow University

ABSTRACT

Speakers pronounce words less clearly when their referents are Given. Speakers' control of intelligibility is shown to reflect a rough, egocentric account of the shared Given information in dialogue. The relative reduction in intelligibility from a first to a second mention of an entity [1] was unaffected by the identity of the original mentioner, the visibility of the entity, and even by the identity of the listener.

INTRODUCTION

Although printed tokens of a word remain uniform, spoken instances will vary, even when they are produced by a single speaker in a single conversation. Part of the typical variability of speech is informative. It is created by the speaker for the listener's sake.

The duration and the intelligibility of different tokens of a word have been shown to be affected by the availability of information other than the token's acoustic shape which might aid word recognition. For example, the more predictable a word is from the sentence context in which it is read, the less intelligible it is (i.e., the lower the proportion of listeners recognizing it) when it is excerpted from that context [2]. More interesting from the point of view of the comprehension of extended discourse, word tokens are less intelligible when they refer to Given entities, both those which are, in Prince's [3] terms, *textually evoked* by previous literal mention [1, 4] and those which are *situationally evoked* by the visible presence of named item [5].

The tendency to degrade redundant tokens seems wonderfully cooperative in Grice's sense of the term[6]: speakers seem to follow a maxim of articulatory quantity in adjusting acoustic information to meet listeners' needs. In the appropriate contexts, less intelligible repeated tokens are actually helpful to listeners, for they make better prompts to earlier discourse material, either because they signal listeners to associate the word's meaning with some entity already established in a discourse model [1, 7] or simply because stored information must be called into play for successful on-line recognition of such items [8].

Unfortunately, degraded tokens are not restricted to contexts in which the listener can recover the conditioning information. Excerpted word intelligibility is equally closely correlated to predictability from sentence context (as assessed by adults) in spontaneous speech to small children and to adults [10], despite the fact that small children have far less mastery of the syntax and vocabulary of those sentences. Word intelligibility reduces across repetitions of a parent's utterance to his/her child, despite the fact that adults repeat themselves to children precisely because children appear not to have noticed earlier tokens of the utterance [5]. Word intelligibility also reduces when adults edit their recorded dictations in the workplace, even though the audio-typist will never hear the original version of the utterance because the speaker has just intentionally erased it [4].

We report several studies designed to determine how far speakers' adjustment of intelligibility is keyed to shared knowledge and how far it uses the speaker's own knowledge to model what is Given for the listener.

EXPERIMENT 1

The first experiment asked how cooperatively speakers interpret the notion 'textually evoked' in dialogue. If speakers adjust their clarity in response to their own contributions, only entities they have introduced themselves are Given and only *self*-repetitions should show a repetition effect, a loss of intelligibility from first to second (co-referential) mention. If speakers contribute to common Given set, however, then both *self*-repetitions and *other*-repetitions should show the effect.

Method

Corpus. To determine what controls the speaker's adjustments we must have an accurate idea of the information which speaker and listener command jointly and individually, including information which each has about what the other knows. For this reason all experiments reported here used word tokens from the HCRC Map Task Corpus [10]. In the 128 unscripted conversations of this corpus, pairs of speakers collaborated to reproduce on one's schematic map a route printed on the other's. Neither speaker could see the other's map. All information relevant to the task appeared on the maps. The speakers' maps differed in to some extent in the names, number, and location of landmarks. Speakers were warned in advance that their maps would not match exactly. Each speaker was Instruction Giver twice for the same map, each time with a different Instruction Follower, though no participant was ever Follower on a particular map more than once. After participating in a series of dialogues, each speaker read a list of landmark names covering the maps just used. (For other details of design see [10].) All materials were recorded on DAT (Sony DTC1000ES) using one Shure SM10A close-talking microphone and one DAT channel per participant.

These design factors make it possible to find word tokens in spontaneous speech which are supported to different degrees by knowledge shared between speaker and listener and to compare their intelligibility with 'citation' or list forms of the same items uttered by the same speaker.

Design and Procedure. The materials were all single word tokens: 48 introductory mentions of shared landmarks, that is landmarks appearing on both maps, second mentions of these by the same speaker, 48 more first mentions by one speaker and their repetitions by the other, as well the same items read by the same speakers in a list.

As in all the other experiments reported below, words were excerpted from digitally recorded materials by examining spectrogram and time-amplitude waveform representations and listening to the results of excerptions. Cut points were set at 0-crossings. Each original word speech file was multiplied, sample by sample, by a 16KHz file of random noise (where all sample values were in the range 0.5 to 1.5) of the same length. In each resulting stimulus, amplitude was related to that of the original speech and data points had the same signs as the sampled data values they replaced. Stimuli were presented from DAT with an ISI of 8 seconds. Word tokens were allocated to different presentation tapes according to a Latin square design.

Eight groups of 10 undergraduate subjects from the same population as the original speakers attempted to identify the words, with only one token of every type heard by any one subject.

Results

Scores for correct recognition were then submitted to ANOVAs by subjects and by materials. ANOVAs were performed both on *raw intelligibility*, the proportion of correct identifications, and *intelligibility loss*, the difference in rate of correct identifications between careful citation tokens and spontaneous mentions. The loss analyses remove differences in baseline citation form intelligibility from consideration. The two kind of analysis conform on the critical results. Table 1 and the reported statistics are based on intelligibility loss.

As Table 1 showed, the effect of repetition, though significant [F_1 (1, 72) = 5.90, $p < .02$; F_2 (1, 80) = 3.26, $p = .075$] was not sensitive to the identity of the original introducer of the entity [Repeater x mention: $F_1 < 1$; $F_2 < 1$, with both first mentions at .15 less than their respective citations forms and both second mentions at .23 less]. Since intelligibility is lost to the same degree for entities which either speaker has introduced, speakers would seem to retain a single common record of textually evoked given entities.

EXPERIMENT 2

The next pair of experiments asked about the applicability of the notion 'situationally evoked', that is, Given by virtue of being present to the senses as mention is made. A critical part of the map task is finding the landmarks mentioned by the other speaker. A cooperative speaker might mitigate intelligibility loss in a second mention if s/he knew that a landmark did not appear on the listener's map. The critical comparison used tokens of items which were textually evoked for both participants, because they had been mentioned before, and situationally evoked for the speaker, who could see them. Only some of these were reported as being present on the listener's map.

Table 1. Difference between intelligibility of citation and running speech forms of words repeated under several conditions (Experiments 1-3)

Stimulus Categories	Mention	
	Token 1	Token 2
Expt 1: Speakers		
Same	.15	.23
Different	.15	.23
Expt 2: Listener can see referent		
Apparently	.24	.16
Apparently not	.13	.21
Expt 3: Speaker can see referent		
Yes	.15	.30
No	.23	.42
Expt 4: Different listeners		
1st pass	.07	
2nd pass	.18	

Method

Sixty first mentions, second mentions, and citation forms were found which named unshared features in two conditions: *Apparently unshared* items were re-iterated after the listener had explicitly denied having the feature; *Apparently shared* items were repeated after the listener erroneously failed to report their absence on his or her map. All 360 word tokens were overlaid with noise and distributed among 6 groups of 9 Subjects each for identification.

Results

Clarity was lost to the same degree when the listener apparently shared the landmark (token 1: .24 less that citation form, token 2 .16 less) and when s/he apparently did not (token 1 .13 less, token 2 .21 less). Though there was an interaction between sharing and mention by subjects [$F_1(1, 48) = 9.05$, $p < .005$; $F_2(1, 96) = 2.32$, n.s.] to which we will return, second tokens did not differ by post hoc tests.

EXPERIMENT 3

Speakers' might have been insensitive to listeners' ability to see landmarks because they lacked interest in what listeners could bring to the task of recognizing words. Alternatively, visual access might affect intelligibility only in creating Given status [5], not in reinforcing it. To test this hypothesis, effects of the *speaker's* own visual access to the named item were examined.

Method

All items were items introduced and repeated by different participants, so that comparison required examing first, second, and respective citation form mentions. For 48 such sets, the repeater did not have the relevant landmark on his/her own map. For another 48, s/he did. Four groups of 9 Subjects were used.

Results

Intelligibility was reduced in one speaker's repetion of another's introduction to the same degree whether (.15) or not (.19) the repeater had visual access to the landmark named. [Mention: Min $F'(1, 116) = 10.52$, $p < .005$; Mention X visual access: $F_1 < 1$; $F_2 < 1$]. Apparently, once an item is textually evoked, neither speaker's access to supplementary visual information will affect delivery.

EXPERIMENT 4

In this experiment we ask whether the set of Given entities are marked with the identity of the individuals who know they are Given. In a map task dialogue, the Giver's strategy ought to be keyed to how much the Follower knows. Givers instruct two different Followers in a single map route. A cooperative Giver should introduce each landmark to the second Follower as clearly as s/he did to the first.

Method

The stimuli here were introductory tokens of the same landmark names uttered by the same speaker in 2 dialogues using the same map but differing in the identity of the listener. Forty-eight item triples were used (first mention to first listener, first mention to second listener, citation form) were used. Because there were some lexical duplications, the 48 were divided into 2 sub-sets and distributed by Latin square among 3 groups of 9 Subjects for identification.

Results

Second-pass introduction (textually evoked for the speaker though New for the new listener) showed greater intelligibility loss vis-a-vis the citation form (.18) than the first-pass introduction (.07) (New for both speaker and listener)(Scheffé at $p < .01$). Once an entity is entered in a representation of material textually-evoked-by-anyone, speakers appear to be insensitive to whether the current listener was witness to a previous mention.

CONCLUSIONS

The general conclusion is that intelligibility is closely controlled by the absence or presence of the named entity in a record of material textually evoked within a dialogue. Once represented in this record, an entity is named by more degraded word tokens, regardless of any other speaker or listener knowledge about the earlier mention or the entity. This arrangement bespeaks a limitation in speakers: although the basic modelling keeps a record of the shared dialogue, modeling listeners minutely while giving accurate instructions may be too burdensome a combination of tasks. In most natural dialogues, where speaker and listener are together in time and space, and where the speaker is not iterating the same message for a succession of listeners, the simplifying assumptions are correct: speaker and listener hear and see the same things, and should remember the same things about a conversation. Hence tracking any differences in situational or textual context is unnecessary. Like those ground-dwelling birds which retrieve the [11] largest visible round object when their eggs roll off the nest, speakers demonstrate what is usually a harmless oversimplification. If, however, there are large round stones near the bird, or if the listener does not share a viewpoint with the speaker, then the error is not harmless at all. The listener may be at as great a disadvantage as the oyster-catcher's egg,

In support this final claim, we can cite subsidiary results from Experiment 2. We had examined repetitions of names of landmarks which appeared on the speaker's map but not on the listener's. Now we looked at the first mentions which preceded the two responses open to the listener, correctly denying having the landmark, and incorrectly failing to report its absence. The first tokens with faulty replies were unusually unintelligible for introductory mentions. Since less intelligible tokens may be interpreted as referring to Given information, we may be dealing here with speakers' egocentric errors that carried a cost.

This work was supported by the ESRC(UK) via the HCRC. Dr. Doherty-Sneddon is now at the Department of Psychology, University of Stirling. *Address for correspondence: E. G. Bard, HCRC, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LL, UK*

REFERENCES

[1] Fowler, C. & Housum, J. (1987), "Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction", *JML, 26*, 489-504.

[2] Lieberman, P. (1963), "Some effects of the semantic and grammatical context on the production and perception of speech", *LangandSpeech, 6*, 172-5.

[3] Prince, E. (1981), "Toward a taxonomy of given-new information", In P. Cole (Ed.), *Radical Pragmatics*, pp. 223-255. New York: Acadmic Press.

[4] Bard, E., Lowe, A., & Altmann, G. (1989), "The effects of repetition on words in recorded dictations", *Proc. EUROSPEECH '89, 2*, 573-6.

[5] Bard, E.G. & Anderson, A. H. (1994), "The unintelligibility of speech to children: Effects of referent availability." *JChLang, 21*, 623-48.

[6] Grice, H.P. (1975). "Logic and conversation." In P. Cole & J. Morgan (Eds), *(Syntax and Semantics: Vol 3, Speech Acts*, pp. 41-58. New York: Academic Press.

[7] Terken, J., and Nooteboom, S. (1987). "Opposite effects of accentuation adn deaccentuation on verification latencies for Given and New information." *LCP* 2, 145-63.

[8] Bard, E. G., Cooper, L., Kowtko, J., and Brew, C. (1991), *Psycholinguistic Studies on the Incremental Recognition of Speech: A Revised and Extended Introduction to the Messy and the Sticky.* ESPRIT Basic Research Action BR3175, DYANA: Dynamic Interpretation of Natural Language, Deliverable R1.3B/C.

[9] Anderson, A. H., et al. (1991). "The H.C.R.C. Map Task Corpus." *LangandSpeech, 34*, 351-366.

[10] Bard, E., & Anderson, A. (1983), "The unintelligibility of speech to children", *JChLang, 10*, 265-92.

[11] Tinbergen, N. (1951). *The Study of Instinct. London: OUP*