CONTRASTIVE INTONATION: SPEAKER- OR LISTENER-DRIVEN?

Constantijn Kaland, Emiel Krahmer & Marc Swerts

Tilburg centre for Communication and Cognition (TiCC), the Netherlands
c.c.l.kaland@uvt.nl; e.j.krahmer@uvt.nl; m.g.j.swerts@uvt.nl;

ABSTRACT

The literature suggests that there are two factors that explain why speakers mark contrastive information: either because it is easy for themselves or because it helps their listeners. The present study investigates whether speakers indeed take their listeners' knowledge into account when prosodically marking contrastive information. A production experiment elicited references to figures (e.g. *blue triangle*) that contrasted with previously mentioned figures (e.g. *red triangle*). Crucially, the previous figure was either described to the same or to a different listener. Results indicate that speakers prosodically mark a contrast more clearly when addressing the same listener.

Keywords: contrastive intonation, speech production, listener adaptation

1. INTRODUCTION

Speakers of Germanic languages such as English or Dutch may use specific intonation patterns to mark information status. Consider the example "Yesterday I saw a blue car, today I saw a red car". Here, the speaker is likely to make the contrastive information (i.e. *red*) more prominent by means of a pitch accent. From the literature it remains unclear which factors drive a contrastive intonation. Does it mainly reflect the information status for the speaker or for the listener?

It has been suggested that a contrastive intonation serves the listener. Levelt [10] points out that listeners use given information as a 'gestalt' (i.e. *car*, in the example above). By modifying just one property (i.e. *red*) listeners can efficiently stick to the gestalt they had in mind instead of creating a new one. Although in the example above it is sufficient to say "...today I saw a red one", speakers mostly repeat the noun [11]. As Levelt [10] argues, a noun is helpful when falling back on a gestalt; i.e. it is easier for listeners to interpret *red car* than *red one*.

Evidence indeed shows that listeners recognize contrastive information faster when it is uttered with the right intonation pattern [13]. Further, contrastive intonation patterns on the sentence level facilitate the recognition of antecedents, even when they are not explicitly mentioned [2] and even when the pattern is heard a day before [5].

As for givenness, [6] and [7] show that speakers account for what their listeners know. That is, information repeated by the speaker is prosodically reduced more when the listener heard the first mention than when the listener did not. As argued by Galati and Brennan [6], articulation processes are guided by a computationally low-cost 'one-bit' model; the listener either heard certain information before or not. Whether these results generalize to contrastive intonation is addressed in this paper.

Although several studies suggest that a contrastive intonation is helpful for listeners, it may be a reflection of only the speakers' perspective on information. According to Chafe [4] speakers indeed use a contrastive intonation even when the listener is not aware of which information is given, for example when Sherlock Holmes utters out of the blue: "The butler did it" (with a pitch accent on *butler*). In this sentence *butler* may contrast with any other suspect (i.e. *gardener*) of which Holmes' is thinking. Chafe [4] calls this "quasi-given" in that givenness of the antecedent only holds from the speaker's point of view. The prosodic marking of *butler* therefore reflects a contrast for the speaker rather than for the listener.

To tackle the issue sketched above, we investigate whether speakers mark a contrast prosodically when givenness of the antecedent applies only to them and not to the listener ("quasi-given"). We manipulate whether contrastive information is uttered to the same listener who heard the previous reference to the antecedent or to a different listener who did not hear the previous reference to the antecedent. On the basis of earlier research we hypothesize that a contrastive intonation is driven both by speaker- and listener-factors. Thus, it is expected that the speaker strengthens the prosodic marking of contrastive

information when addressing the same listener, as compared to when addressing a different listener.

2. METHOD

To elicit references to contrastive information, participants act as speakers in a referential communication task in which they instruct two different listeners to put figures on a piece of paper. The order of instructions is manipulated so that two successive instructions refer to two figures that can be distinguished by just their colour or just their shape (test stimuli) or by both their colour and their shape (fillers). A test stimulus concerns the latter of two successive instructions, as the present study investigates contrastive intonation with respect to the previous utterance. Two successive instructions are either uttered to the same listener or to different listeners (*listener*: same, different). The setup ensures that only successive instructions to the same listener make sense in terms of contrastive intonation, not successive instructions to different listeners. That is, speakers are told that when addressing one listener, the other listener hears music via a headphone so that the instruction cannot be heard. In reality, listeners are confederates and hear all instructions (see section 2.3). Because contrastive information in the test stimuli concerns either the colour or shape of the target figure, the focused word is either the adjective or noun (*focus*: adjective, noun).

2.1. Participants

20 different participants acted as speaker (17 women, 3 men, M_{age} = 21.8 years, age range: 18-29 years). They were all Dutch speaking students of Tilburg University participating for course credit.

2.2. Design and materials

The communication task is played as a bingo game with the speaker as the game leader and listeners as players. Each listener has a different bingo card displaying 24 common objects (e.g. fruit, tools, means of transport). Bingo cards are 6 x 4 grids with rows numbered from 1 to 4 and columns marked by each character of "bingo!" (Figure 1). In addition, listeners each have a set of paper card figures; a drop, clover, canoe or triangle (in Dutch *druppel*, *klaver*, *kano* and *driehoek* respectively) coloured red, yellow, green or blue (in Dutch *rood, geel, groen* and *blauw* respectively). Different rounds are played, which begin by the speaker's

announcement of which row or column has to be covered by target figures (for example a figure on each cell of row 2). The listener who achieves the right pattern first shouts "bingo!", upon which that listener receives a point and the round ends. The speaker has to keep the scores. The first instruction of each new round is a filler to account for speakers' pitch reset upon switching discourse contexts [3]. The stimulus order occurs in two randomizations; each of which is presented to 10 participants. Speakers utter 48 instructions in total (equally spread over listeners, crossed for the factors listener and focus).

Figure 1: Example of the speaker's screen, showing in Dutch *Beschrijf aan A* (describe to A), the target figure (bottom left) and A's bingo card. A typical instruction would be: "put the red clover on the flag".
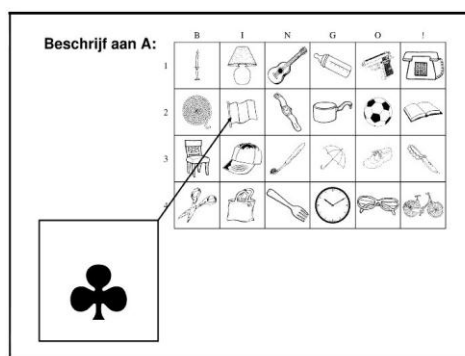

2.3. Procedure

The speaker is seated at one end of a table and listeners, who cannot see each other but who are both visible to the speaker, at the other end (Figure 2). Before the game begins speakers receive instructions and play a training round. Listeners wear open-ear headphones to facilitate the speaker's illusion that the listener who is not addressed hears music. After each experiment speakers are asked whether they indeed believe that listeners heard music and not the instruction (all responded affirmative).

Speakers (not listeners) see a screen displaying the target figure and the bingo card of the listener to be addressed (Figure 1). The screen's lay-out indicates when speakers have to switch between listeners. That is, for listener A the target figure is displayed on the screen's left side and for listener B the target figure is displayed on the right side. In accordance, speakers have to look past the left side of the screen when addressing listener A and past the right side of the screen when addressing listener B (Figure 2). Additionally, speakers are

told that the software responsible for the instruction slides on the screen also switches music between listeners. Speakers' speech is digitally recorded by a headset microphone and saved as wave-file.

Figure 2: Birdseye view of the experimental setup showing the speaker facing the screen (bottom) and the listeners, at opposite sides of a partition, facing their bingo cards and figures (top).
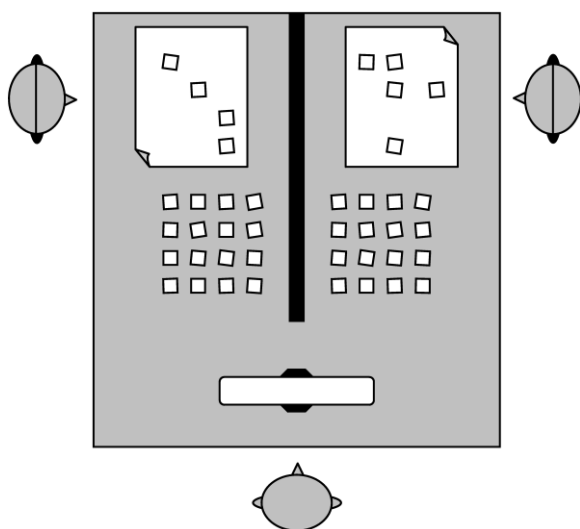
2.4. Prosodic analysis

NPs referring to target figures in the test stimuli (n = 480) were extracted from the wave-file recordings using Praat [1]. They were acoustically analysed in terms of prominence by perception ratings and pitch measures (F0), the latter was taken as a strong correlate of prominence [9]. As for the ratings, NPs were presented in a web-based task [12] to three intonation experts. They rated the strength of the accent on a three point scale (0 = no accent, 1 = weak accent, 2 = strong accent). Adjectives were rated in the first part of the task, nouns were rated in the second part. The presentation order of NPs was randomized so that experts were blind for condition. To abstract over the experts' ratings, the prominence scores per word were added up so that they range from 0 to 6 (0 when all experts rate the accent as absent, 6 when all experts rate the accent as strong). Pearson's correlation coefficients as computed for the adjective and noun ratings indicate that the experts' ratings are consistent [r(478) range = .62 - .72, $p < .01$].

As for pitch measures, F0 maxima in Hertz on the stressed syllable of the adjective and noun were measured in Praat [1]. Some speakers ended the NP with a high boundary tone on the noun's last

syllable. However, that syllable was the never stressed one (see section 2.2).

As shown by [8], the contrastively focused word in Dutch obtains prominence by *both* its accentuation *and* deaccentuation of the unfocused word. To account for this finding a difference score is computed. That is, the prominence score of the unfocused word is subtracted from the prominence score of the focused word. In this way, positive difference scores indicate that the focused word is more prominent than the unfocused word and negative scores indicate that the unfocused word is more prominent than the focused word. The same procedure is carried out for the F0 maxima.

2.5. Statistical analysis

Analyses of Variance (ANOVAs) are performed on repeated prominence and F0 difference scores as dependent variables with listener (2 levels: same, different) and focus (2 levels: adjective, noun) as within-subject factors.

3. RESULTS

As for the prominence difference scores, no negative means are found (Table 1), revealing that overall the focused word is perceived as more prominent than the unfocused word. As for the factor listener, prominence difference scores are larger when the same listener is addressed (M = 2.89) than when a different listener is addressed (M = 1.95): [$F(1,19) = 16.48$, $p < .001$, $\eta_p^2 = .46$]. Further, the difference between the focused word and the unfocused word is larger when the focused word is the adjective (M = 3.51) than when the focused word is the noun (M = 1.33): [$F(1,19) = 11.81$, $p < .01$, $\eta_p^2 = .38$]. How the difference scores relate to the adjective and noun becomes clear from their individual prominence scores. These reveal that both the focused word is less prominent and the unfocused word is more prominent when the listener is different than when the listener is the same (Table 1). Concerning pitch, no main effects of listener or focus are found. However, there is an interaction between the two factors in that addressing the same listener results in larger difference scores for a focused adjective, whereas addressing a different listener results in larger difference scores for a focused noun: [$F(1,19) = 7.21$, $p < .05$, $\eta_p^2 = .28$]. Further, prominence ratings and F0 maxima correlate: [$r_{adjective}(478)$ =.25, $p < .01$] and [$r_{noun}(478) = .10$, $p < .05$].

Table 1: Mean prominence score, mean F0 maximum (Hz) and standard deviation for adjective, noun and their difference as a function of listener and focus.

Listener	Focus	Prominence score M (SD)			F0 maximum M (SD)		
		Adjective	Noun	Difference	Adjective	Noun	Difference
Same	Adjective	5.21 (*1.00*)	1.32 (*1.44*)	3.98 (*1.48*)	340.64 (*138.34*)	279.24 (*117.88*)	61.40 (*160.33*)
	Noun	2.42 (*1.96*)	4.30 (*1.83*)	1.88 (*2.50*)	295.87 (*129.53*)	298.33 (*120.52*)	2.45 (*178.56*)
Different	Adjective	4.79 (*1.43*)	1.67 (*1.63*)	3.13 (*1.87*)	323.15 (*128.55*)	299.80 (*116.70*)	23.35 (*165.54*)
	Noun	2.94 (*1.84*)	3.71 (*1.92*)	0.76 (*1.96*)	254.96 (*64.94*)	302.95 (*120.98*)	47.99 (*125.66*)

4. DISCUSSION

Results show that the prosodic marking of contrastive information is both speaker- and listener-driven. That is, when addressing a different listener speakers still distinguish the contrastively focused word from the unfocused word by means of prosodic marking. This is in accordance with Chafe [4]. However, speakers use a clearer contrastive intonation when addressing the same listener. The latter finding indicates that speakers account to some extent for whether the listener heard the introduction of a 'gestalt' in the previous utterance [8, 10]. If not, a contrast does not have to be as clearly marked as when the listeners can make use of gestalt information.

Inspection of the individual prominence scores of the adjective and noun indicates that when a speaker addresses the same listener, the focused word becomes more prominent and the unfocused word becomes less prominent compared to when the speaker addresses a different listener. As for the unfocused word this outcome is a replication of what is found by [6] and [7] in that given information is reduced more when addressing the same listener than when addressing a different listener. In general, the stronger the inverse prominence relationship of adjective and noun, the clearer the contrastive intonation. This finding confirms that Dutch contrastive intonation depends on both the accentuation of the focused word and the deaccentuation of the unfocused word [8].

Furthermore, it matters what the focused word is. A contrastive adjective is perceived as much more prominent than the non-contrastive noun, whereas a contrastive noun is only moderately more prominent than the non-contrastive adjective. Such a finding is in accordance with [8]. The present results do not tell to what extent this effect is related to (a combination of) prosodic properties, phrase position or nature of the two word classes.

In short, the current study favours the view that contrastive intonation is a conditional optimum of speaker- and listener-factors. It seems as if demands of both interlocutors, even if conflictuous, are taken into account for the production of prosody.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Boersma, P., Weenink, D. 2010. Praat (version 5.1.25). *http://www.praat.org/*

[2] Braun, B., Tagliapietra, L. 2010. The role of contrastive intonation contours in the retrieval of contextual alternatives. *Proc. Language and Cog.* 25, 1024-1043.

[3] Brown, G., Currie, K., Kenworthy, J. 1980. *Questions of Intonation*. London: Croom Helm.

[4] Chafe, W.L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, C. (ed.), *Subject and Topic.* New York: Ac. Press, 25-55.

[5] Fraundorf, S.H., Watson, D.G., Benjamin, A.S. 2010. Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *JML* 63, 367-386.

[6] Galati, A., Brennan, S.E. 2010. Attenuating information in spoken communication: For the speaker, or for the addressee? *JML* 61, 35-51.

[7] Gregory, M.L., Jurafsky, D., Healy, A.F. 2001. *The role of the Hearer in Durational Shortening*. Poster presented at *AMLaP 7* Saarbrücken, Germany.

[8] Krahmer, E., Swerts, M. 2001. On the alleged existence of contrastive accents. *Speech Comm.* 34, 391-405.

[9] Ladd, D.R. 2008. *Intonational Phonology* (2nd ed.). Cambridge; New York: Cambridge University Press.

[10] Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation*. Cambridge, Mass.: MIT Press.

[11] Pechmann, T. 1984. *Überspezifizierung und Betonung in Referentieller Kommunikation*. Unpublished Dissertation, Universität Mannheim, Mannheim.

[12] Veenker, T.J.G. 2003. *WWStim: A CGI script for presenting web-based questionnaires and experiments* (Version 1.4.4). Utrecht University.

[13] Weber, A., Braun, B., Crocker, M.W. 2006. Finding referents in time: eye-tracking evidence for the role of contrastive accents. *Language and Speech* 49, 367-392.