IS IT IMPORTANT FOR COMMUNICATION *WHICH* PARAMETERS SIGNAL ACCENTUATION?

William Barry & Bistra Andreeva

Institut für Phonetik, Universität des Saarlandes, Germany
wbarry@coli.uni-saarland.de; andreeva@coli.uni-saarland.de

ABSTRACT

Two experiments examine the *perceptual* consequences of systematic differences between Bulgarian and German in the *production* of phrasal prominence. In experiment one, five versions of a German sentence differing in focus, and therefore in the degree of prominence of two critical words implicated in the focus conditions, were presented in pairs to Bulgarian and German subjects who judged the *degree* to which the prominence of the critical words differed. In a second experiment, the questions used to elicit the different focus realizations of the sentence were paired with each of the focus versions. Subjects judged how well the answer matched the question. In the meta-linguistic first task, evidence was found to support the hypothesis that the greater use of signal intensity in Bulgarian prominence production is paralleled by greater sensitivity to intensity differences in perception. In the functional second task, there was no indication that the greater sensitivity to signal intensity has any communicational consequences.

Keywords: accentuation, perception, cross-language differences

1. INTRODUCTION

1.1. Segmental definition

There is ample evidence – ranging from reduction and elision of syllables and words (e.g. [6]) to the replacement of a word by pink noise [10] – that the precise segmental skeleton of an utterance is not exclusively important for utterance comprehension (and is certainly not the only contributor to communicative ease). Residual and contextual phonetic information is used, and with the support of a prosodic structure (supporting 'perceptual coherence') reflecting the information structure of the intended utterance, the semantic content that would have been carried by the segments is reconstituted.

As the extreme example of [ə̰̃ə̃ə̰̃] for "*I don't know*" in [4] illustrates, the leeway for *segmental* distortion is very considerable. The leeway that is allowed for *prosodic* distortion, on the other hand, has not often been considered, and it is certainly not assumed that prosodic ambiguity is compensated for by articulatory accuracy, although there are examples of segmental properties affecting prosodic meaning. There has long been evidence that durational manipulation of segments can effect focus interpretation [5] and the co-variation of fricative noise frequency and intonation contour to extend the pattern obscured by the voiceless fricative has been reported recently [7].

1.2. Accentuation and language differences

The perceptual correlate of placing a linguistic 'accent' on a word in order to bring it into focus is the increased prominence of the stressed syllable of the word. Conversely, a de-accented word has reduced prominence. Of the four acoustic parameters that are universally available to induce an increase or decrease in perceived prominence – duration, intensity, f0 and spectral energy distribution – it has been shown that different languages use them to differing degrees, i.e. they are differently weighted in production from language to language [1, 8]. With increasing levels of accentuation, Bulgarian (Bg) and German (G), for example, show different degrees of change in the three main accent-bearing parameters – f0, intensity and duration. G employs increased duration significantly more than Bg, which shows less durational change with increased emphasis. Conversely, Bg increases syllable intensity with accentuation significantly more than G. F0, on the other hand is clearly important in both Bg and G.

Recent work [3] has shown that in perception too, languages attach different degrees of importance to duration and f0 in functional prosodic judgments. The two parameters interact in marking group boundaries. Swiss German speakers attach a lesser degree of importance to increased syllabic duration and greater importance to a rising f0 contour than (Swiss) French speakers when deciding where a sequence of five elements is divided into two groups (e.g. 12+345 or 123+45).

The goal of this study is to consider the phonetic detail of phrasal accentuation for information-structural purposes (focus) across languages. It seeks to determine whether the degree of inter-language variation that has been observed in the use of duration, intensity and f0 in the *production* of focused and non-focused lexical items has *perceptual* consequences for the communicationally important interpretation of focus

2. EXPERIMENTS

2.1. Speech material

G and Bg were chosen because analyses of their prominence-giving production patterns had revealed the differences reported above.

A G sentence “Der Mann fuhr den Wagen vor” (The man brought the car round) was used in the experiment; the two ‘critical words’ (CW) under scrutiny were CW1 “Mann” and CW2 “Wagen”. The five questions presented to the speakers in the original production experiments (Andreeva et al. 2007) are shown in table 1.

Table 1: German questions used for eliciting utterances with broad and (non-contrastive and contrastive) narrow focus questions.

Question	**Focus condition**
Was passierte? (*What happened?*)	Broad
Wer fuhr den Wagen vor? (***Who** brought the car round?*)	Non-contr. early(CW1)
Was fuhr der Mann vor? (***What** did the man bring round?*)	Non-contr. late (CW2)
Die **Dame** fuhr den Wagen vor? (*The **lady** brought the car round?*)	Contr. early (CW1)
Der Mann fuhr die **Klagen** vor? (*The man brought the **charges** round?*)	Contr. late (CW2)

These questions triggered a broad-focus response, a non-contrastive and a contrastive narrow-focus response for CW1, and a non-contrastive and a contrastive narrow-focus response for CW2.

For experiment 1 each focus condition was paired with itself (identical pairing) and with each of the other four conditions, in both orders of presentation, giving 45 pairs with either identical or various differing degrees of prominence for each CW. A part of each version of the sentence was masked by low-pass filtering to help the subjects concentrate on the relevant CW. “Der Mann” was masked when “Wagen” was to be judged; “fuhr den Wagen vor” was masked when “Mann” was to be judged.

The five identical pairings were used to ascertain the basic ability of the subjects to judge relative prominence. Five repetitions of each stimulus pair in each order of presentation were offered to the groups, giving 225 stimulus pairs in total. Five practice pairs, which were not evaluated, preceded the stimuli for familiarization with the task.

In experiment 2, the recording of each of the 5 questions used to elicit the different focus conditions of the test sentence was paired with a realization of each of the answers. This resulted in 5 question-answer (QA) pairs with the original matching realizations and four pairs each of every other combination (e.g. Q broad + A non-contrastive early, contrastive early, non-contrastive late, contrastive late), a total of 25 stimulus pairs. Again, 5 repetitions were offered, giving 125 stimuli, presented in randomized order.

2.2. Experimental tasks and subjects

In experiment 1, subjects judged greater or less prominence; experiment 2 was a communication-linked task dependent on the ability to judge the information structure of an utterance.

In experiment 1, Bg and G subjects were presented with G sentence pairs which differed only in the focus conditions under which the sentences were produced. The task was to judge to what *degree* a critical word in one version differed in prominence from the same word in the other. Subjects registered their judgments on a graphic interface by moving a slide upwards or downwards on an uncalibrated scale (with concealed values of ±100). The zero-difference position was marked in the centre of the scale. Moving the slide upwards signified a more prominent CW in the first sentence of the pair. Moving it downwards signified a more prominent CW in the second sentence.

Experiment 2 used the same five focal versions paired with different focus-determining questions. Subjects judged on a scale of 1-5 how well the version offered fitted the question, 1 indicating an optimal fit, 5 indicating that the question and answer did not fit together at all. Subjects' decision times were recorded at the same time.

30 Bg and 30 G subjects judged the same G sentences (Bg_g and G_g). The comparison of these two response sets is central to the hypotheses to be tested, but in experiment 2, Bg subjects also

judged Bg Q-A pairs (Bg_bg) to indicate whether prosodic marking of information structure (IS) is equally important in Bg and G.

2.3. Response-data pre-processing

The uncalibrated scale used in exp. 1 allowed each subject to set his/her own level of response. Since some subjects make fuller use than others of the scale's range, averaging subjects repeat responses can distort the differences between groups. Therefore z-values for each subject were used to normalize for individual scaling differences.

The stimulus pairs were allocated to four stimulus groups: 1. identical stimuli, 2. 'near-identical' stimuli with the same narrow-focus location (early or late) but differing in level (non-contrastive vs. contrastive), 3. stimulus pairs comprising the broad-focus realization with a narrow-focus version (early or late, non-contrastive or contrastive), 4. maximally different stimulus pairs comprising *early* narrow-focus sentences with *late* narrow-focus ones.

In exp. 2 the combinations of Q and A are grouped into three categories:

1. Q paired with the actual A it elicited (expected response 1 or 2 = "good match").
2. Q eliciting a broad-focus A paired with a narrow-focus A, or Q eliciting a narrow-focus A with a broad-focus A. (expected response uncertain).
3. Q eliciting an *early* narrow-focus A paired with a *late* narrow-focus A, or a Q eliciting a *late* narrow-focus A paired with an *early* narrow-focus A (expected response 4 or 5 = "bad match").

3. RESULTS

3.1. Experiment 1: prominence-difference

The z-values of the individual means for each stimulus pairing were used in a univariate ANOVA with the subject group (Bg and G) and the stimulus group (1-4) as factors. Overall, the subject groups were not significantly different (p=0,263), while the stimulus group, as expected, differed highly significantly ($p<0.001$). Post hoc tests showed that all four stimulus groups differed significantly from one another. There was also a significant interaction between the subject groups and the stimulus groups, which warranted closer scrutiny.

Bg and G show almost identical behaviour for the identical pairs (p=0.962) and were not significantly different for the 'near-identical' pairs (p=0.504) nor for the pairs containing broad-focus stimuli (p=0.172). However, they differed significantly for the '*maximally different*' stimulus pairs (p=0.003).

Examination of this stimulus group revealed that the difference between the Bg and G subjects was entirely due to a divergence in their prominence difference judgment of *contrastive early* focus realizations paired with prenuclear deaccented CW1 in a) the non-contrastive (= diff 1) or b) contrastive (= diff 2) *late* focus condition. For these stimulus pairs the Bg subjects registered a 12.5% *greater* prominence difference than the G subjects.

Considering the acoustic differences between the CW1 realizations against the background of the general Bg and G production patterns [1, 8], the following implications for the subjects' perceptual behaviour can be identified:

- the differences in *tonal* movement on "Mann" (diff 1: 8.2, diff 2: 8.6 ST) should signal similar prominence differences for both groups;
- diff 1: 60.2, diff. 2: 46.4% *durational* differences should be important for the G but less so for the Bg subjects,
- diff 1: 3.8, diff. 2: 4.8 dB differences in *intensity* should contribute to the perceived prominence difference for the Bg but not for the G subjects.

This should, it would seem, result in similar overall judgments for Bg and G subjects, duration and intensity differences compensating for each other across the subject groups. But an explanation for the 12.5% greater difference for the Bg subjects lies in the psycho-acoustic integration of signal energy and duration which operates for durations up to 250-300ms [9]. So, both G and Bg subjects react to the durational difference, though for different reasons, whereas only the Bg subjects react to the increase in intensity. Thus, they perceive the same *signal* differences as greater *prominence* differences.

3.2. Experiment 2: question-answer matching

Table 2 summarizes the responses of the three subject groups in exp. 2. A multivariate ANOVA with the subject groups and the QA-categories as factors was carried out with the subjects' mean acceptability judgments and their mean decision times for each QA combination as dependent measures. The QA-categories differed significantly, as expected: $p<0.001$), but subject groups differed also:

p<0.001. Post hoc test showed that all three subject groups differed significantly in their acceptability judgments.

Table 2: Absolute and percentage distribution of judgments for the three QA categories for the Bg_bg, Bg_g and G_g subjects.

Subject group	Re-sponse	1	2	3	4	5
Bg_bg	Cat.1	922	246	25	115	42
	%	68.3	18.2	1.9	8.5	3.1
	Cat.2	306	387	84	272	151
	%	25.5	32.2	7.0	22.7	12.6
	Cat.3	79	88	49	447	537
	%	6.6	7.3	4.1	37.3	44.7
Bg_g	Cat.1	1020	198	29	49	54
	%	75.6	14.7	2.1	3.6	4.0
	Cat.2	354	425	74	252	45
	%	29.5	35.4	6.2	21.0	3.8
	Cat.3	97	111	36	392	564
	%	8.1	9.3	3.0	32.7	47.0
G_g	Cat.1	1126	194	10	16	4
	%	83.4	14.4	0.7	1.2	0.3
	Cat.2	138	483	11	378	90
	%	11.5	40.3	19.3	31.5	7.5
	Cat.3	2	9	18	315	856
	%	0.2	0.8	1.5	26.2	71.3

The G subjects were most nearly categorical and most clearly symmetrical in their acceptance of matching QA-combinations (cat. 1) and their rejection of contrasting QA-combinations (cat. 3), with 97.8% acceptance and 97.5% rejection (vs. 86.5% acceptance/82% rejection for Bg_bg and 90.3 acceptance/79.7% rejection for BG_g). Also, the G_g decision times were significantly shorter than both Bg groups (G_g: 5.6s < Bg_bg: 6.9s = Bg_g: 7.2s; p<0.001). Another difference between the G subjects and both Bg groups is their rating of cat. 2 combinations. Whereas 81.1% of the G_g responses fall in categories 2-4 (= "don't know" or weak acceptance /rejection), more than 80% of the Bg responses are in categories 1, 2 and 4 (Bg_bg 80.4% with 57.7% acceptance vs. 22.7% weak rejection; Bg_g 85.9%, with 64.9% acceptance vs. 21% weak rejection).

The parallel perceptual behaviour of the Bg_bg and Bg_g groups along these various dimensions, in particular the strong acceptance of prosodically contrasting QA-focus-pairings, suggests strongly that the Bg subjects rely less heavily on the prosodic differentiation of IS than the G subjects. This assumption is borne out by [2]. Avgustinova models IS in Bg utterances as an interplay of three factors: the lexeme-specific obliqueness hierarchy of grammatical relations, the observable constituent order, and the contingent clitic replication. While prosodic factors are also included under the general term “intonation”, one may assume that there is considerably more grammatical signalling for the Bg than for the G listener. Our data show that Bg subjects do not de-accentuate and do not register de-accentuation in communicative terms. Most important for the question posed in this study: There is no evidence that the differences in Bg and G evaluation of information structure is in any way influenced by their differential sensitivity to signal intensity.

4. CONCLUSIONS

The conclusion to be drawn from these experiments is that production differences between languages may have perceptual-processing consequences in a non-functional discrimination task, but that such differences do not necessarily carry over to a more communication-linked functional task.

5. ACKNOWLEDGMENTS

This research was funded by the German Research Council, grant BA 737/10-1.

6. REFERENCES

[1] Andreeva, B., Barry W., Steiner I. 2007. Producing phrasal prominence in German. *Proc. 16th of ICPhS* Saarbrücke, 1209-1212.

[2] Avgustinova, T. 1998. Word order and clitics in Bulgarian. *Saarbrücken Dissertations in Computational Linguistics and Language Technology* 5. Universität des Saarlandes DFKI.

[3] Cumming, R.E. 2010. *Speech Rhythm. The Language Specific Integration of Pitch and Duration.* PhD. Dissertation, University of Cambridge, UK.

[4] Hawkins, S., Smith, R. 2001. Polysp: a polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics - Rivista di Linguistica* 13, 99-188.

[5] Huggins, A.W.F. 1972. Just noticeable differences for segment duration in natural speech. *Journal of the Acoustical Society of America* 51, 1270-1278.

[6] Kohler, K.J. 1990. Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In Hardcastle, W.J., Marchal, A. (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 69-92.

[7] Kohler, K.J. 2011. To be appeared. Communicative functions integrate segments in prosodies and prosodies in segments. *Phonetica*.

[8] Koreman, J., Andreeva B, Barry, W. 2008. Accentuation cues in French and German. *Proc. Speech Prosody*, 613-616.

[9] Moore, B.C.J. 2003. *An Introduction to the Psychology of Hearing*. Academic Press.

[10] Warren, R.M., Obusek, C.J. 1971. Speech perception and phonemic restorations. *Perception & Psychophysics* 9, 358-362.