A Cross-language Corpus for Studying the Phonetics and Phonology of Prominence

Bistra Andreeva[1], William Barry[1], Jacques Koreman[2]
[1]Saarland University – Germany
[2]Norwegian University of Science and Technology – Norway
andreeva@coli.uni-saarland.de, wbarry@coli.uni-saarland.de, jacques.koreman@ntnu.no

Abstract

The present article describes a corpus which was collected for the cross-language comparison of prominence. In the data analysis, the acoustic-phonetic properties of words spoken with two different levels of accentuation (de-accented and nuclear accented in non-contrastive narrow-focus) are examined in question-answer elicited sentences and iterative imitations (on the syllable 'da') produced by Bulgarian, Russian, French, German and Norwegian speakers (3 male and 3 female per language). Normalized parameter values allow a comparison of the properties employed in differentiating the two levels of accentuation. Across the five languages there are systematic differences in the degree to which duration, *f0*, intensity and spectral vowel definition change with changing prominence under different focus conditions. The link with phonological differences between the languages is discussed.

Keywords: prominence, acoustic correlates, cross-language

1. Motivation for corpus collection

Prominence is a perceptual phenomenon which, in spoken language, results from the acoustic realisation of phonological structuring at two levels: lexical stress and phrasal accentuation within the utterance. The phonetic basis of prominence has long been accepted as comprising the relative duration, *f0* (difference or movement), intensity and spectral properties of the (vocalic) unit. Excitation quality as a possible correlate of prominence (cf. Koreman, 1996, Marasek, 1997) is related to both intensity and spectral quality, and has rarely been considered separately. Of the traditional parameters, duration and *f0* have been shown experimentally to be more important in perceived prominence in English than intensity and degree of spectral reduction (Fry, 1955, 1958, 1965). However, the contribution of *f0* to prominence has not been borne out in analyses of large speech databases (cf. Van Kuijk & Boves, 1999; Kochanski et al., 2005). In agreement with received wisdom, the simple dB measure of syllable strength did not prove important, but more refined measures of signal energy suggest a revision of earlier assumptions. Van Kuijk & Boves (1999) found that either a combined value of intensity and duration, duration alone, or a spectral tilt measure performed best in classifying (lexically) stressed and unstressed vowels separately for each vowel phoneme. In linguistically carefully controlled data, Sluijter & Van Heuven (1996) and Sluijter et al. (1997) also found that spectral tilt is a valid acoustic and perceptual correlate of stress and accent. Kochanski et al. found that their acoustic "loudness" measure (based on Stevens 1971) was the primary correlate of accentuation, more important even than duration. In contrast, Streefkerk et al. (1999), using the same database as van Kuijk & Boves, found that the traditionally more important parameters *f0* range and duration were the best predictors of perceived prominence. These studies, however, all had binarily labelled databases (± prominent auditory judgments / ± lexical stress derived from the lexicon) rather than differentiated judgments of greater or lesser prominence to base their analyses on. We note discrepant results between the studies, but cannot say whether they stem from differences in material (full-band vs. telephone-quality speech), differences in the language material (Dutch vs. English) or different approaches to the analysis (auditorily judged prominence vs. lexical stress).

Within a linguistic framework, cross-language comparisons are logically the primary frame in which to seek (a) how different languages exploit the universal (= psycho-acoustically determined) means of modifying the prominence of words in an utterance; (b) whether the different word-phonological requirements of a language affect the degree to which the properties are exploited, and (c) whether differences between languages are greater than the differences between speakers of a language. We are NOT investigating "word stress/word accent", but rather the change in a given word as a result of making it more or less prominent in the utterance by varying the information structure.

For the analysis of the exploitation of the four accepted stress/accent determining acoustic properties (duration, *f0*, intensity and vowel spectrum), a corpus of read speech was recorded.

2. Languages and Speakers

The languages covered in the corpus are assumed to belong to different "rhythm types" and also differ in basic phonological properties: variable vs. fixed word stress (or lacking word stress), presence vs. absence of a vowel length distinction, variable vs. low syllable complexity; in addition, the languages differed in their phonological and phonetic reduction mechanisms in unstressed syllables. The following languages were recorded: German as a northwestern European "stress-timed" language, Russian as "stress-timed" and Bulgarian as "syllable-timed"

Slavonic languages with different vowel reduction patterns, French as a clear "syllable-timed" candidate, since it has no lexical stress, Japanese as a mora-timed language and Norwegian as a language which is not so readily categorized as either stress- or syllable-timed.

Six tertiary-educated, regionally homogeneous speakers (3 male and 3 female) per language were recorded: for German, speakers from the Saarland area who spoke Standard German; for French, speakers of northern standard French; for Bulgarian, speakers of Sofia-Bulgarian; for Russian, speakers of Standard Russian from the Moscow area; the Japanese informants were all speakers of Tokyo Japanese and the Norwegian informants were all speakers of Urban East Norwegian. The regional homogeneity aimed at increasing the chance of a group hierarchy in the exploitation of the acoustic dimensions, i.e. the regional sub-stratum which could have influenced the establishment of their prominence-giving mechanisms was constant. This design choice also implies that the results found in this study may not be directly generalizable to other variants of the languages investigated.

3. Material and recordings

In order to provide a basis for the direct comparison of parameter values across different conditions of phrasal accentuation in the five languages, controlled utterances with a canonical word order were required for each language which could be produced with de-accented and accented variants of the same words. We believe that a laboratory corpus, made up of several "artificial" utterances created specifically for the task provides more reliable data than a spontaneous speech corpus, since it permits the isolation of the variables under study as well as the neutralisation of other factors. Six short sentences per language were constructed containing two one- or two-syllable "critical words" (CWs), one early (but not initial) and one late (but not final) in the sentence. For each sentence, a number of questions were devised to elicit a) a broad focus response, b) a response with a non-contrastive narrow focus on the early CW1 and c) on the late CW2 and d) a contrastive focus on the early CW1 and e) on the late CW2.

4. Recordings

The speakers produced 6 repetitions of each of the sentences from a PowerPoint presentation in response to the pre-recorded questions in a sound-treated studio. To provide a basis for comparing the parameter modification across sentences independently of the different segmental structuring of the critical words (and thus, if possible, to derive a speaker- and/or language-specific quantification of the accent-dependent modification), a reiterative 'dada' version of each realisation was produced immediately after the normal-text response. This was produced in two stages: (i) a 'da' or 'dada' replacement of only the (mono- or disyllabic) CWs and (ii) a 'da' replacement of all the syllables in the sentences.

The recordings were made using an AKG C420IIIPP headset on a Tascam DA-P1 DAT recorder and transferred digitally via the optical channel to a PC using the Kay Elemetrics MultiSpeech speech signal processing program. The corpus consists of totally 19440 sentences (6 languages x 6 speakers x 6 sentences x 5 focus conditions x 6 repetitions x 3 versions).

5. Labelling

Segmentation, labelling with slightly modified SAMPA transcriptions and further processing were done using the Kiel XASSP speech signal analysis package. Special labels are used to refer to sub-phonemic events like closure and release of the stops, devoiced portions of voiced segments and vice versa, etc. (cf. Figure 1). Six labelling assistants were allocated different sentences (to maximize labelling consistency across conditions within each sentence) and segmentation problems were regularly discussed and decided with the authors at group level.

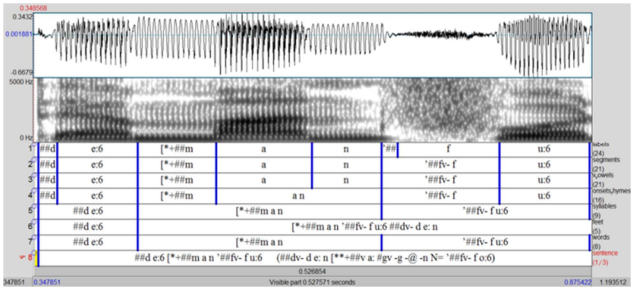

Figure 1: Example of the segmental labelling for the first part of the sentence 'Der Mann fuhr den Wagen vor' (The man brought the car round).

6. Measurements

Parameters in the four acoustic dimensions were calculated using praat scripts and operationalized as follows:

(a) *Durations* were calculated for all feet in the sentences, for the CWs and their component syllables as well as the syllables of the feet to which the CWs belonged. Furthermore, the duration of the phonetic sound segments comprised in the syllables were calculated. All durational measurements were normalized as a percentage of the mean duration of the corresponding unit in the sentence.

(b) Since comparisons focus on changes in identical words across conditions, *f0* was calculated as the mean fundamental frequency (Hz) across the syllable nucleus (vowel or syllabic sonorant) of the lexically stressed syllable of CWs and in the unstressed syllable preceding and following it. The average *f0* across the utterance was subtracted to normalize the *f0* values.

(c) *Intensity* was measured in two ways: first, as the mean intensity (in dB) of the stressed vowel in the CW, and second, as the spectral balance in that vowel. This was computed as the energy difference between the frequency

band from 70-1000 Hz and that from 1200-5000 Hz. This measure, too, was normalized by subtracting the spectral balance across the whole utterance.

(d) *Spectral definition* was captured with the mean frequency (and bandwidth) values for the first three formants in the middle of the syllabic nuclei in the lexically stressed syllable of CWs.

Several analysis methods were applied to the acoustic parameters. For the sake of optimal comparability across languages, most of the analyses were carried out on the reiterant 'dada' utterances, although the results were always verified for the text (replies to eliciting questions).

7. Results

In the analyses presented in this article, only a subset of the data is studied. For all languages except Japanese, only responses to the questions eliciting a *non-contrastive narrow focus* a) on the early CW1 and b) on the late CW2 were investigated. This leads to two degrees of prominence on the critical words: (a) de-accented and (b) nuclear accented (cf. Figure 2).

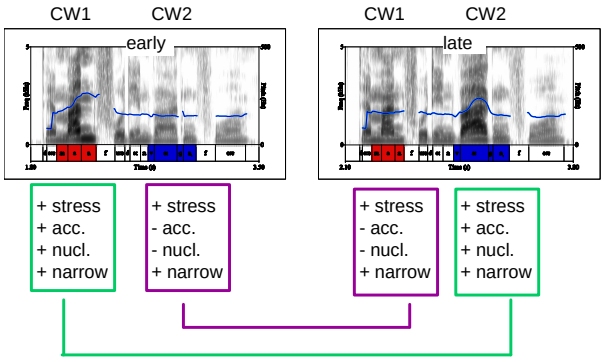

Figure 2: Level of prominence on the CWs

As a first step towards specifying the differences between the accenting and de-accenting patterns in Bulgarian, German, French, Norwegian and Russian, one-way repeated measures ANOVA's per parameter were carried out for CW1 and CW2 separately, with accent level (accented, de-accented) as a within-subject and language (BG, G, F, N, RUS) as a between-subject variable. We report univariate tests with Greenhouse-Geisser estimates of F. Separate Tukey post-hoc tests were carried out per variable, if appropriate. The confidence level was set at α=0.05. We present results for the 'dada' material since this allows direct comparison across the five languages without distortion from different syllable structures. Table 1 shows the main effects for language over different degrees of accentuation for CW1 and CW2.

These main effects indicate that the five languages behave differently with regard to their normalized duration and *f0* values, their spectral tilt and in the values of the second formant.

Parameter	CW1	CW2
syllable duration	*	n.s.
onset duration	n.s.	*
vowel duration	*	n.s.
f0 mean	**	*
f0 change	***	**
Intensity	n.s.	n.s.
spectral tilt	*	*
F1	n.s.	n.s.
F2	*	n.s.
F3	n.s.	n.s.

Table 1: Main effects for language (BG, G, F, N, RUS) (* p<0.05; ** p<0.01; *** p<0.001).

While this is important to register, the link between these differences and the accenting and de-accenting process in speech production is only indirect – namely, by virtue of the fact that the data reflect the mean and variance of the parameters in the CWs, which have been produced in a context defined as "de-accented" and "accented". More important for the issue addressed in this study are the interactions between language and accent level. These reflect the parameters that are exploited differently in the process of accenting and de-accenting.

Table 2 shows the interactions for CW1 and CW2. For CW1, and even more clearly for CW2, the five languages differ significantly in the degree to which they employ duration for accent differentiation.

Parameter	CW1	CW2
syllable duration	*	***
onset duration	n.s.	*
vowel duration	*	***
f0 mean	n.s.	n.s.
f0 change	**	*
intensity	**	**
spectral tilt	n.s.	**
F1	n.s.	n.s.
F2	n.s.	n.s.
F3	n.s.	n.s.

Table 2: Interactions for language (BG, D, F, N, RUS) x degree of accent (* p<0.05; ** p<0.01; *** p<0.001)

Across the five languages (Bulgarian, Russian, German, Norwegian and French) there are systematic differences in the degree to which duration, intensity, *f0* and spectral vowel definition change with changing prominence. Table 3 and Table 4 show the significant language-group differences for CW1 and CW2, respectively.

For CW1, the increase in syllable duration with accentuation is greater for Norwegian than for Bulgarian ($F[4, 25] = 3.216$, $p < 0.05$). The increase in vowel

duration with accentuation is greater for Norwegian and French than Bulgarian ($F[4, 25] = 3.604$, $p < 0.05$). Russian and French differ significantly in the degree to which they employ *f0* change for prominence differentiation ($F[4, 25] = 4.479$, $p < 0.01$). Bulgarian and Norwegian differ significantly in the degree to which they employ intensity for prominence differentiation ($F[4, 25] = 4.310$, $p < 0.01$). There is no difference in the use of spectral balance and *f0* mean values despite the main language effects. The languages also do not differ in the manner in which the spectral definition of the vowel (the change in the quality as reflected in the formant values) changes between the de-accented and the accented condition (cf. Table 3).

syl. dur.	N = F = G = R > F = G = R = B
vowel dur.	N = F = R = G > R = G = B
f0 change	F = G = B = N > G = B = N = R
intensity	B = F = G = R > F = G = R = N

Table 3: Significant language-group differences for CW1.

For CW2, Norwegian utilizes syllable duration more strongly than Russian, German and Bulgarian, and French utilizes syllable duration more strongly than Bulgarian ($F[4, 25] = 10.362$, $p < 0.001$). Norwegian exploits vowel onset durational change for accentuation purposes to a considerably greater degree than Russian ($F[4, 25] = 3.003$, $p < 0.05$). Norwegian increases the vowel duration more strongly than German and Bulgarian ($F[4, 25] = 10.172$, $p < 0.001$). As far as *f0* is concerned French and German use *f0* change stronger then Norwegian ($F[4, 25] = 3.346$, $p < 0.05$). Bulgarian, French and German use intensity significantly stronger than Norwegian ($F[4, 25] = 5.353$, $p < 0.001$). French and German on the one hand and Russian and Norwegian on the other differ significantly in the degree to which they employ spectral tilt ($F[4, 25] = 5.651$, $p < 0.01$) (cf. Table 4).

syl. dur.	N = F > F = R = D > R = D = B
onset dur.	N = F = D = B > F = D = B = R
vowel dur.	N = F = R > F = R = D > R = D = B
f0 change	F = D = B = R > B = R = N
intensity	B = F = D = R > R = N
spec. tilt	N = R = B > R = B = D > B = D = F

Table 4: Significant language-group differences for CW2.

8. Conclusion and Discussion

The findings of this study confirm the primary hypothesis that languages will differ systematically in the degree to which they employ the four acoustic dimensions underlying different levels of phrasal accentuation. The differences between the languages in relation to their exploitation of the acoustic dimensions duration, *f0*, intensity and spectral definition can be only partially explained with reference to differences in the phonological structure of the five languages.

The picture that emerges from the production results with regard to the phonologies of the languages involved is not at all clear in its implications. Clear cases of accenting strategies which conform to expectations derived from phonological patterning stand alongside opposing strategies.

If we assume that the lack of a vowel-length opposition provides ‘space’ for greater accentual lengthening and shortening for information-structural purposes, then French exploits that space while Bulgarian does not. Intriguingly, the two Germanic languages behave differently in the late sentence position. An obvious language-intrinsic explanation does not present itself. For example, none of these languages is constrained in its phrasal prosody by any use of duration for lexical stress purposes, as is the case e.g. for Italian.

Intensity changes with degree of accentuation differ little across languages except for Norwegian, which exploits it to a lesser degree than Bulgarian in early sentence position and than Bulgarian, French and German in late sentence position. The relative importance of intensity carries little or no phonological implication, with the exception of the possible importance attached to the expression "dynamic accent" in relation to Slavonic languages. In that respect, the results for Bulgarian confirm the expectation of a greater intensity range in the critical words across the focus conditions; less so for Russian.

Spectral balance, which is recognized as a correlate of the unstressed-stressed distinction at lexical level, but which has not previously been examined in relation to degree of phrasal accentuation, is shown to vary most strongly in Norwegian. Not surprisingly, given the lack of lexical stress marking in French, French speakers show a less systematic use of spectral balance change with accentuation. The use of *f0* change with accentuation is relatively strong across all languages.

Behind the statistically highly significant language differences, individual speakers were found to diverge in one or another parameter from the dominant pattern of the language (Andreeva et al., 2007; Koreman et al., 2009). Against this background of individual freedom within language differences, it is perhaps unsurprising that the relationship of the exploited prominence-giving properties to the phonologies of the languages is not clear-cut. Their use sometimes conforms to, sometimes diverges from Dauer’s hypothesis (1983, 1987) that properties exploited at the lexical level are not available at the phrasal level, and vice versa. Most strikingly, duration conforms in the case of French and German, but not in Bulgarian, Russian and Norwegian. Fundamental frequency use conforms more generally, being constrained by lexical tone in Norwegian, but not in the other languages. However, the manner in which it is used, namely the change of tone accent with change of focus or

just increased range with increased prominence, varies across languages. A detailed intonation analysis is clearly necessary, and the parameterization of *f0* difference may need refinement to capture not only the degree but also the type of change. The result then may well be a number of differences in prominence-giving production behaviour linked to the intonational phonology of the languages.

9. Acknowledgements

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

10. References

Andreeva, B., Barry, W.J. and Steiner, I. (2007). Producing Phrasal Prominence in German. *Proc. ICPhS*, Saarbrücken, pp. 1209--1212.

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *J. Phonetics* 11, pp. 51--62.

Dauer, R. (1987). Phonetic and phonological components of language rhythm. *Proc. XIth ICPhS*, vol. 5: 447-450.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress, *J. Acoustical Soc. America* 27, pp. 765--8.

Fry, D. B. (1958). Experiments in the perception of stress, *Language and Speech* 1, pp. 126--152

Fry, D. B. (1965). The dependence of stress judgments on vowel formant structure, *Proc. VIth ICPhS*, pp. 306--311, Basel: Karger.

Kochanski, G., Grabe, E., Coleman, J. (2005). Loudness predicts prominence: fundamental frequency lends little. *J. Acoust. Soc. America* 118, pp. 1038--1054.

Koreman, J. (1996). *Decoding Linguistic Information in the Glottal Airflow*. Dissertation, Nijmegen University.

Koreman, J., Andreeva, B., Barry, W.J., van Dommelen, W. and Sikveland, R.-O. (2009). Cross-language differences in the production of phrasal prominence in Norwegian and German, In: Martti Vainio, Reijo Aulanko, and Olli Aaltonen (eds.), *Nordic Prosody, Proceedings of the Xth Conference*, Helsinki 2008, Frankfurt: Peter Lang, pp. 139--150.

Kuijk, D. van & Boves, L (1999). Acoustic characteristics of lexical stress in continuous telephone speech, *Speech Communication* 27, pp. 95--112.

Marasek, K. (1997). *Electroglottographic Description of Voice Quality.* (= AIMS 3,2, Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung, Lehrstuhl für experimentelle Phonetik), Universität Stutgart.

Sluijter, A., van Heuven, V. (1996). Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. America* 100, pp. 2471--2485.

Sluijter, A. M. C., van Heuven, V.J. & Pacilly, J. J. A.(1997). Spectreal balance as a cue in the perception of linguistic stress, *J. Acoustical Soc. America* 101, pp. 2471--2485.

Stevens, S. S. (1971). Perceived level of noise by Mark VII and decibels, *J. Acoustical Soc. America* 51, pp. 575--602.

Streefkerk, B. M., Pols, L. C. W. and ten Bosch, L. F. M. (1999). Acoustical features as predictors for prominence in read aloud Dutch sentences used in ANN's, *Proc. EUROSPEECH'99*, Vol. 1, pp. 551--554.