

Prosody perception in L1 and L2

Irina Nesterenko

¹ Institut National des Langues et Civilisations Orientales (INALCO)

irina_nesterenko@yahoo.com

Abstract

The present work deals with the perception of prosodic structure in conversational speech by naïve listeners in L1 and L2. Native speakers of Russian and French-speaking subjects with high proficiency in Russian took part in a prosody transcription experiment. After evaluating inter-transcriber agreement, we analyse the interplays between the distributions of perceived prosodic prominences and boundaries and discuss the influences of prosodic patterns specific to listeners' native language.

Index Terms: Prosody perception, Prosody transcription, Spontaneous Speech, L2.

1. Introduction

The structuring of speech material in terms of boundaries and groupings is one of the main functions of prosody. Within such prosodic domains, some elements are more salient than their neighbours – one speaks then about prosodic prominences' distribution. Both prosodic phrasing and prosodic prominences determine the prosodic organisation of speech which highlights the hierarchical organisation of a message and its division in terms of old / new information. As a consequence, prosodic organisation is key in showing the link among syntax, semantics, discourse and information structure. It follows that to recover prosodic organisation from the speech signal is of importance for the listener to understand the utterance: a body of psycholinguistic research has shown that this information is of importance for effective speech processing and ambiguity resolution [2, 7].

One of the methods to investigate the perception of prosody is prosodic transcription studies. Previous work investigated the perception of prosodic boundaries and prominences by experts and naïve transcribers in both read-aloud and spontaneous speech [1, 3, 6, 8, 9]. Considerably high inter-transcriber agreement rates were observed, especially with simplified inventories of prosodic categories.

We chose to apply the method of prosodic transcription to the issue of prosody perception in L2. We investigated the prosody perception in casual speech as this is the most frequent context in which non-native speech communication take place. In fact, the ability to recover information about the prosodic structure of an utterance, and thus about its syntactic, semantic and information structure is of even greater importance for efficient communication in L2. In our study, we investigated how naïve French speaking subjects with high proficiency in Russian perceive prosody of Russian connected discourse, thus addressing the issue of whether the L2 speakers interpret prosodic organisation in a manner similar to L1 speakers. After describing the experimental design, we present the data on inter-transcriber agreement and on the structural properties of perceived units. Finally, we put the observed patterns of perceived prosodic structure into the perspective of structural differences between French and Russian (cf. the absence of lexical stress and high correlation

between prominence and boundary phenomena in French) and discuss the influences of prosodic patterns from the listeners' native language.

2. Method

2.1. Materials

Our study is based on a 10-minute excerpt from the corpus of spontaneous dialogue speech in Russian, collected for the INTAS project 915 at the department of Phonetics, Saint-Petersburg State University. For the current study, the stimuli were selected from the recordings of an informal spontaneous dialogue between two female speakers in their twenties. The selected material was divided in 20 sound files, respecting the division into Inter-Pausal Units. In the perceptual study a printed version (in Russian) of the recorded material was provided to the subjects. The transcripts did not contain any punctuation marks or capitalisation; on the contrary, speech errors and disfluencies were signalled in the transcripts.

2.2. Procedure

Five Russian-speaking subjects and five French-speaking subjects with high proficiency in Russian took part in the experiment. They are members of the Mutualised Seminar on Formal Slavic Linguistics of Paris IV / INALCO universities. The participants are faculty members, PhD students or graduate students in linguistics with no knowledge of the autosegmental-metrical theory of prosodic organisation. In a brief introduction we presented the goals of our study and proposed quite general definitions of prosodic grouping (the perceived breaks and groupings in the word string) and prominence (any salient element standing out from its neighbours). We then instructed our participants to underline prominent words and mark prosodic breaks by vertical lines. All the sound files were played three times, the subjects annotating as they listen. The subjects were sitting in a classroom and had no graphical representation of the speech signal: the resulting transcriptions are thus perception based.

The data were tabulated in an Excel file, with every word marked for the presence of a prosodic boundary after it and the perception of its prominence by each listener. The total of 331 words is analysed. In the following analyses, we evaluated inter-transcriber agreement rates, the differences in the perception of prosodic organisation as a function of differences in the agreement rates and we looked at the interplay between prosodic grouping perception and the distribution of prosodic prominences, analysing the co-variation between the two annotations and the distribution of prosodically salient words within the identified prosodic groups.

3. Results

3.1. Inter-transcribers' agreement tests

On average, Russian speaking and French speaking listeners marked the presence of prosodic break every 2.5 words (min = 2.06, max = 3.31). Prosodic prominence was perceived on every third word (min = 2.07, max = 5.06).

To evaluate inter-transcriber agreement we had recourse to Fleiss kappa statistics [3] presented in Table 1: Fleiss statistics provide a single measure of agreement between all pairs of transcribers and in this respect differ from mean Cohen's kappa score for all the pairs of judges used in original studies on prosodic transcription [8 and references cited there].

Z-scores in Table 1 are all statistically significant ($p < 0,001$). The values of Fleiss Kappa do not differ much when the data received in both groups are pooled together and when the data for L1 and L2 subjects are analysed separately. As in previous studies of naïve prosody transcription in spontaneous speech in L1 (cf. [2] for American English, [7] for French), we observe higher agreement for boundary perception than for prominence judgements.

We further analysed the variation in agreement rates between all the pairs of transcribers using Cohen's kappa coefficients. The distributions of Cohen's kappa coefficients are presented in Figure 1 for boundaries and in Figure 2 for prominences. Kappa scores in our experiment ranged from 0.576 to 0.767 for boundary annotation (for Russian-speaking subjects min = 0.579, max = 0.767; for French-speaking subjects min = 0.576, max = 0.749) and from 0.248 to 0.622 for prominence judgements (for Russian-speaking subjects min = 0.297, max = 0.622; for French-speaking subjects min = 0.248, max = 0.562).

Globally, L2 listeners' performance is comparable with that of native speakers.

Table 1: *Fleiss Kappa scores.*

	All subjects	L1 subjects	L2 subjects
Fleiss K for boundaries	0,645 $z=78,7$	0,658 $z=37,8$	0,646 $z=37,2$
Fleiss K for prominences	0,469 $z=57,2$	0,469 $z=27$	0,493 $z=28,3$

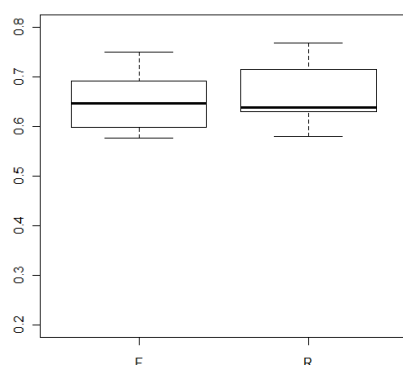


Figure 1: *Cohen's Kappa coefficient distribution in the perception of prosodic grouping.*

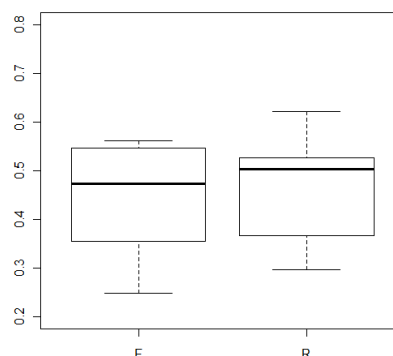


Figure 2: *Cohen's Kappa coefficient distribution in the perception of prosodic prominence.*

3.2. Interplay between prosodic grouping and prominence perception

Next, we analyzed the interplay between prosodic phrasing and prosodic prominence perception. The traditional view on Russian prosody stipulates that there is one prosodic prominent element per unit, though in informal speech there could be patterns of prosodic grouping with two prosodically highlighted elements.

Figures 3 and 4 plot the counts of perceived prosodic groups against the number of prominent elements identified within such a group. Three patterns were distinguished: a) prosodic groups without any prominent element, b) prosodic groups with 1 prominent element; c) prosodic groups with 2 and more perceptually prominent elements. All the above mentioned configurations are found for both L1 subjects and L2 subjects.

On average, L1 subjects identify 60.5% (min=57%, max = 73.3%, sd = 8.72%) of units with one prosodic prominence, 30% of units with no prominence (min = 20.9%, max = 43.1%, sd = 8.98%) and 9.5% (min = 3.42%, max = 14.9%, sd = 4.39%) of units with two and more prosodic prominences. L2 subjects marked 56% (min=51.5%, max = 63.3%, sd = 7.48%) of units with one prosodic prominence, 22.75% of units with no prominence (min = 16.31%, max = 32.26%, sd = 4.92%) and 21.17% (min = 15%, max = 28.79%, sd = 5.94%) of units with two and more prosodic prominences.

As to the differences between the two groups of listeners, proportion tests show that only the differences in number of units with two prominences is statistically significant at the 95% level ($\chi^2 = 6,43$, $df = 1$, $p = 0,011$): L2 subjects annotated more units with two prominences. This could be the consequence of subjects' native language: in French, as opposed to Russian, there is no lexical stress; as a result, subjects probably promoted lexical stress to the level of phrasal prominence. At the same time, the existence of

units with two and more elements suggests that, though the construction of prosodic domains starts with metrical units, the variance in relative strength of prominence, in melodic and temporal information allows the construction of larger prosodic groups. We notice a significant inter-annotator variability both for the zero prominence group and the two and more prominences group.

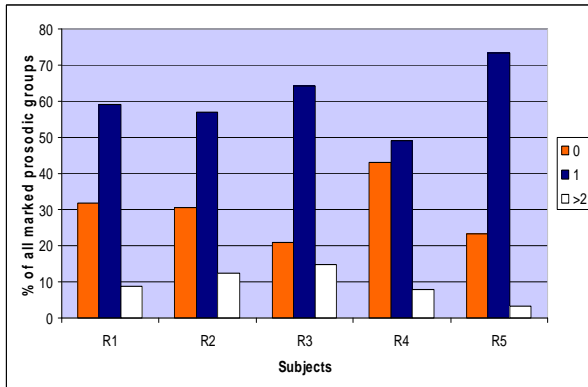


Figure 3: Number of perceived prosodic prominences per prosodic unit, Russian-speaking subjects

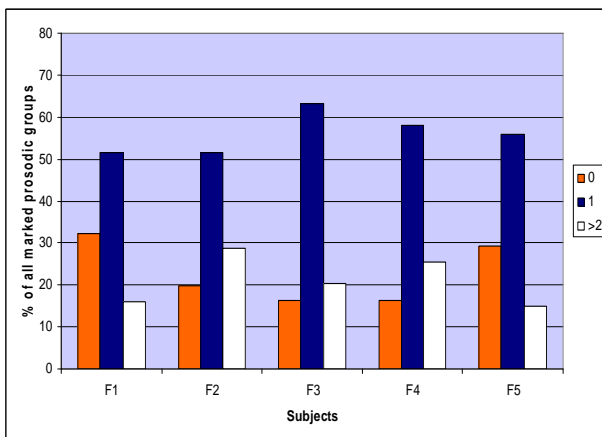


Figure 4: Number of perceived prosodic prominences per prosodic unit, French-speaking subjects

3.3. Probabilistic Break and Prominence Indexes

In our further analysis we followed the methodology proposed by [2], which assigns to each orthographic word a probabilistic index, corresponding to the proportion of listeners who marked this word as prominent or as preceding a prosodic boundary. These probabilistic Break indices and Prominence indices were used to evaluate the correlation between these prosodic phenomena.

Correlation tests revealed a strong correlation between prosodic prominence and prosodic break perception for both L1 (cor = 0.568, $t = 12.5113$, $df = 329$, $p < 0.001$) and L2 subjects (cor = 0.585, $t = 13.08$, $df = 329$, $p < 0.001$) at the 95% confidence level. It means that when listening to Russian spontaneous speech both Russian-speaking and French-speaking participants tend to mark a prosodic boundary in the vicinity of a prosodic prominence. A similar

result was obtained by [8] for French speaking subjects listening to French spontaneous speech.

4. Discussion

This study looks at the perception of prosodic organisation of conversational speech by L1 and L2 listeners. Results on inter-transcriber agreement rate show that naïve listeners are consistent when labelling prosodic boundaries and prosodic prominences both in L1 and in L2. Participants were more consistent in annotating prosodic groupings than prominences. We note as well a significant variation in the perception of prosodic phenomena across different pairs of listeners, which is however comparable between L1 and L2 groups.

While analysing the distribution of prosodic prominences within prosodic groups, we find that there is one prosodic prominence annotated in about 60% of units. At the same time, while Russian-speaking subjects perceive two or more prominent elements in only 9% of units, this score increases to 21% in the L2 group. When comparing these two patterns of responses, it seems that French-speaking subjects promote some lexically stressed elements to the level of phrasal prominence. On the other hand, 30% of units marked by L1 listeners and 22% of units in L2 listeners' responses had no prominence. A brief survey of units with no prominence shows that this category regroups, among others, some parenthetical expressions (traditionally such elements form a prosodic unit of their own and are characterised by a suppressed pitch range) and units corresponding to the de-accented post-focal stretches of speech in narrow focus contexts. In this respect, further investigations should examine whether a focused element systematically triggers the perception of prosodic boundary and the strength of this boundary (whether the resulting unit is an Intonational or an Intermediate phrase).

Lastly, the analysis of probabilistic Boundary and Prominence scores reveals a significant interplay between prominence and boundary perception both for L1 and L2 subjects. In fact, models of prosodic organisation for French [4] do insist on a strong link between prosodic prominences and prosodic boundaries. This link is weaker for such languages as English, in which from a theoretical point of view there is no particular constraint on the location of prominence with respect to the boundaries, except in cases of neutral declaratives: in this latter case, the word bearing a nuclear accent is usually the last content word in the utterance. Russian, being a language with lexical stress, seems to us to more closely resemble English, than French. Our result may thus have at least three implications for further research : a) the performance of French participants reveals that the strategy of associating prosodic prominence with boundary phenomena was transferred when analyzing prosody in L2; b) we need to further investigate the issue of prosodic phrasing in narrow focus utterances with special emphasis put on the status of prosodic break after the focused constituent from a typological perspective; c) finally, a large scale statistical estimation from prosodically annotated databases could reveal statistical dependencies between the distributions of prosodic prominence and prosodic boundary.

5. References

- [1] Buhmann J., Caspers, J., van Heuven, V, Hoekstra, H., Martens, J.-P. & Swerts, M. 2002. Annotation of Prominent Words, Prosodic Boundaries and Segmental Lengthening by Non Expert Transcribers in the Spoken Dutch Corpus. *Proc. LREC 2002*, 779-785.
- [2] Clifton, C. Jr., Carlson, K., Frazier, L. 2002. Informative prosodic boundaries, *Language and Speech*, 45:87-114.
- [3] Cole, J., Mo, Y., Baek, S. 2010. The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, 25:7, 1141-1177.
- [4] Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 76, 3780382.
- [5] Jun, S.-A., Fougeron, C. 2002. Realizations of accentual phrase in French intonation. *Probus* 14, 147-172
- [6] Mo, Y., Cole, J., Lee, E. 2008. Naive listeners' prominence and boundary perception. *Proc. Speech Prosody 2008*, 735-738.
- [7] Price, P., Ostendorf, M., Shattuck-Hufnagel, S., Fong, C/ 1991. The use of prosody in syntactic disambiguation, *J. Acoust. Soc. Am*, 90:2956-2970.
- [8] Smith, C. 2009. Naïve listeners' perception of French prosody compared to the predictions of theoretical models. *Proc. IDP 09*, 335-349.
- [9] Syrdal, A.K., McGory, J. 2000. Inter-transcriber reliability of ToBI prosodic labelling, *Proc. ICSLP*, vol. 3, 235-238.