



# Some evidence on the phonetics and phonology of prosodic phrasing in Russian

Irina Nesterenko<sup>1</sup>, Pavel Skrelin<sup>2</sup>

<sup>1</sup>Laboratoire Parole et Langage, Université de Provence, France  
<sup>2</sup>Department of Phonetics, Saint-Petersburg State University, Russia  
 irina.nesterenko@lpl.univ-aix.fr, skrelin@phonetics.spb.ru

## Abstract

This paper treats the issue of prosodic segmentation into phrasing domains in Russian and is framed in the prosodic phonology paradigm. Distributions of prosodic boundaries are obtained in a perception experiment and the results are further explored to advance the hypotheses about the levels of prosodic constituency in Russian. The temporal organisation of the perceived domains and the eurhythmic constraints on phrasing are investigated. Particularly, the empirical data suggest that beyond the level of intonational units, there are two other levels, a level of metrical domain and one of phonological phrase, relevant in the perception of phrasing patterns.

**Index Terms:** Russian, prosodic phrasing, phonological constraints

## 1. Introduction

This study deals with the issue of prosodic phrasing in Russian. The structuring of speech material in terms of boundaries and groupings is one of the main functions of prosody. Prosodic phrasing has been the subject of advanced formal modelling within the framework of prosodic phonology [1-2] and a motivated hierarchical structure of prosodic constituents has been proposed for many languages. It should be underlined that such a representation allows the distinction between different boundary strengths, what crucially influences the quality of synthesized speech in the applied studies [3]. Concurrently, a body of processing studies have shown that the way the words are grouped together to form larger linguistic units affects the understanding of the utterances [4]. Yet, we do not know of any study conducted on Russian within the framework of prosodic phonology.

When discussing the structuring function of prosody, Russian prosodic research refers to the concept of syntagma [5]. This unit is defined as a “minimal intonational unit which can be coterminous with syntactic constituents; it can also be viewed as a cognitive unit formed in the process of speaking”. As follows from the definition above, several principles, crucial to the prosodic phonology approach, are shared by the Russian prosodic tradition, namely the independency and complex relations between prosodic and syntactic structures of the utterance and the role of the Sense Unit Condition [6] in defining higher level prosodic constituents. Yet, in empirical Russian studies of prosody the principles underlying prosodic segmentation are not clearly defined and the researchers rely mainly on syntactic information in segmenting the speech flow into syntagmas.

In the framework of prosodic phonology, a highly developed theory of prosodic domains and prosodic structure was proposed for several European languages (English,

French, Italian or Dutch) but its transfer to other languages is less clear since terminological and conceptual differences make the interpretation of the findings difficult. As the literature analysis reveals, the syntagma of Russian prosodic studies could be compared to an accentual phrase, phonological phrase, intermediate phrase or intonation unit, as those units are defined in various phonological studies. In the rest of our paper we present the results of a perception study designed to collect the data on the distribution of prosodic boundaries in Russian spontaneous speech and to advance hypotheses about levels of prosodic phrasing in Russian. It has been demonstrated that the place of prosodic boundaries is influenced by such factors as information structure and eurhythmic constraints together with speech rate. While discussing the levels of prosodic constituency in Russian, we test the relevance of different eurhythmic constraints.

## 2. Methodology

Our study is based on 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. 25 inter-pausal units (IPU) of variable length and syntactic complexity were selected; table 1 resumes several temporal characteristics of the stimuli.

Table 1. *Temporal organization of stimuli*

	Mean	Standard Deviation	Min	Max
Duration (s)	2.87	1.12	1.34	5.28
Length (syllables)	17.6	5.8	6	30
Length (words)	7.2	2.4	3	12

To set up an experimental study in order to collect data on the distribution of prosodic boundaries is a notoriously difficult task since it amounts to eliciting meta-linguistic judgements. Even if a population of experts is solicited, there is considerable influence of their previous theoretical background as well as of the training received. We also kept in mind the fact that a prosodic boundary is a complex acoustic phenomenon and that the acoustic cues for boundaries may vary across different languages.

This motivated our recourse to a cross-linguistic study: two groups of trained phoneticians, native speakers of French and Russian, took part in the experiment. Moreover, to ensure that the listeners do not rely exclusively on the syntactic and segmental information, the stimuli were filtered. Consequently, the proposed task of annotating the distribution of prosodic boundaries was realised under three experimental conditions:

- **CondF** (for *flat*): the stimuli were presented with flattened fundamental frequency (the flattening being obtained via PSOLA resynthesis with the value set to the mean F0 value for the stimulus); afterwards, a low pass filter was applied (threshold fixed at 500 Hz).
- **CondD** (for *delexicalised*): a delexicalised version of the stimuli was obtained by application of a low-pass filter (same threshold).
- **CondN** (for *normal*): a natural sounding version was presented during the last session.

This manipulation allowed us to obtain unintelligible stimuli with a rather good quality of preserved prosodic information. We expected as well that to work with three experimental conditions would allow us to collect the data about the progressive building of the prosodic hierarchy.

7 Russian-speaking subjects and their 7 non Russian-speaking (in our case, French speaking) counterparts took part in the experiment: all subjects were PhD students in the field of phonetics/prosody or faculty members of the speech labs.

The experiment was thus composed of three sessions for each subject. The task proposed to the subjects consisted in marking the intonational phrase boundaries, without any explicit definition of the concept being proposed. Each subject fulfilled the task individually. The listeners were asked to make judgments on the basis of acoustic cues only. They could listen to the speech stimulus or its parts as many times as they wished. The subjects navigated through the speech file displayed via the Praat program with a blank tier reserved for them to enter their boundary marks.

Before proceeding with the phonetic analyses, we ascertained that there was sufficient agreement across the listeners. To do this, we had recourse to the kappa statistic (K, [7]): the highest agreement characterises the performance of Russian speakers under non-manipulated condition (K = 0.76). Otherwise, the performance of French speaking subjects and that of Russian speaking subjects under the manipulated conditions are comparable and exhibit kappa values around 0.5. We judged these agreement data sufficiently good to proceed with further analyses.

### 3. Results

The perception experiment allowed us to collect the data about the distribution of prosodic boundaries. These data were further analysed to validate the relevance of different prosodic constituents for Russian. Particularly, we place an Intonational Unit at the highest level of the prosodic hierarchy: being associated with a pitch contour, this unit is universally acknowledged in the prosodic studies. At the same time, the experimental data suggest that another two prosodic constituents are relevant for the analysis of Russian prosody, namely, phonological phrases and rhythmic units. These conclusions are supported by the examination of the following parameters: number of boundaries annotated under different experimental condition, phonological length of the perceived units (in number of syllables) and their duration (in ms).

#### 3.1. Number of prosodic boundaries perceived

To evaluate the impact of the Condition factor on the number of prosodic boundaries perceived, we had recourse to a mixed

linear model. The responses of Russian speaking and French speaking subjects were analysed separately.

The anova of the linear model (NumberOfBoundaries ~ Condition) reveals a significant effect of the Condition factor in the performance of Russian speaking subjects ( $F(2, 348) = 5,836$ ,  $p = 0.0032$ ), this effect being absent in the performance of French speaking subjects ( $F(2, 348) = 1.089$ ,  $p = 0.3377$ ). The further analysis of fixed effects reveals that there is a significant difference in the number of annotated boundaries between CondF, on the one hand, and the CondD and CondN, on the other hand for Russian participants (cf. table 2): more boundaries are signalled when there is no pitch information available. Note that for the parameter analysed, there is no difference in the number of boundaries marked under CondD and CondN by Russian speaking subjects, nor for French speaking subjects.

Table 2. Fixed effects of the number of boundaries as a function of experimental condition (Russian subjects)

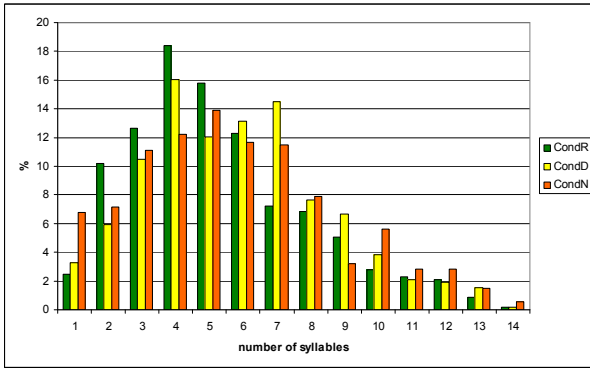
	t value	p value
CondF vs. CondD	-2.922	0.0037
CondF vs. CondN	-2.99	0.0030
CondD vs. CondN	-0.071	0.9432

Our results clearly show that Russian speaking subjects annotate significantly more prosodic boundaries under the flattened f0 condition than when pitch information is available (CondD and CondN). This tendency underlines the role of the intonation in the perception of phrasing. At the same time, as its formal properties, the syntagma is traditionally defined in Russian prosodic studies as a speech chunk bearing a coherent pitch contour. Under the CondF condition, all the pitch information has been removed, but all the rhythmic and intensity information was preserved: the metrically stressed syllable could thus be interpreted as centre of prominence and the rhythmical unit promoted in the prosodic constituency to the level of Intonation Unit. The data are compatible with the hypothesis of the progressive formation of prosodic units: grouping starts with metrical domains and next, pitch cues assure the formation of larger units, which belong to higher levels of prosodic constituency.

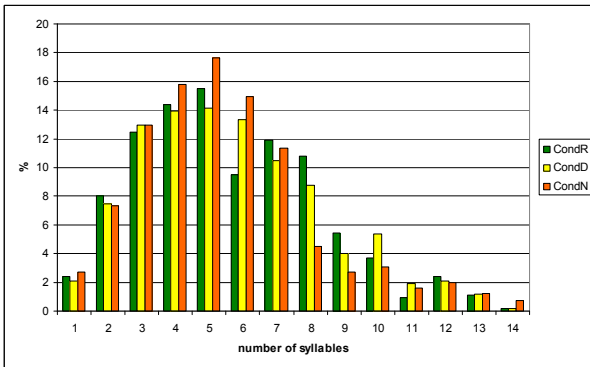
#### 3.2. Phonological length

We further analysed the rhythmic properties of perceived prosodic units via the parameter of phonological length (expressed in number of syllables). The importance of eurhythmic considerations for the prosodic structuring of speech has been underlined in numerous studies [8, 9]. Moreover, we suppose the existence of a rhythmical model of prosodic units as part of the linguistic competence, explorable under the manipulated conditions.

We established that the mean length of the prosodic units in our experiment is 5.73 syllables for Russian speaking subjects (standard deviation 2.998 syllables) and 5.77 syllables for French speaking subjects (standard deviation 2.98), the values being rather stable between experimental conditions. We also observed a high value for the coefficient of variation (around 0.5): this means that the data exhibit considerable dispersion around the mean value.



**Figure 1:** Frequency distribution of prosodic units as function of their phonological length (Russian subjects)



**Figure 2:** Frequency distribution of prosodic units as function of their phonological length (French subjects)

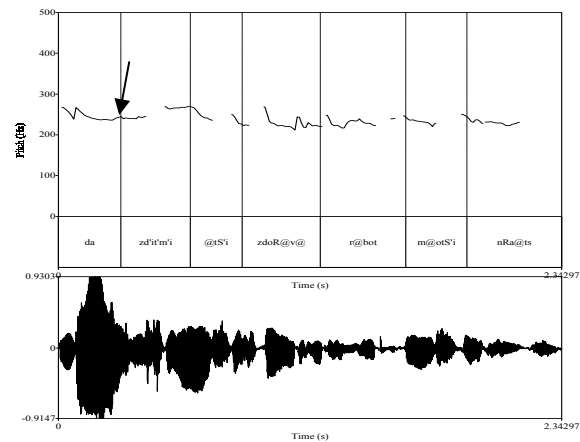
Figures 1 and 2 present the frequency distribution for the units of different length. These data were analysed with reference to the phonological constraints [10]. First, we determined the proportion of units containing less than 3 syllables (table 3). Note that these units are more frequent in the annotation of Russian speaking subjects under normal and flattened  $f_0$  conditions. At the same time, they mark twice as many monosyllabic units under the normal condition as compared to the other two conditions. This result reflects the tendency to place discourse markers (“yes” and “no” particles) and introductory words in separate intonation units. Certainly, a special status of parenthetical expressions or other sentence-external elements is underlined in many prosodic studies. The question is whether the boundaries of such elements are sufficiently motivated by prosodic cues (durational or tonal) in spontaneous speech: figure 3 illustrates that sometimes there is no marked prosodic boundary (accompanied by a boundary tone) after such an element.

A further critical limit is set at the level of 6 syllables. In our data (table 3), around 63% of the perceived units contain 6 syllables or less. The length of the prosodic word in Russian fluctuates between 2.75 and 3.16 syllables, with the general mean set at 3 syllables [11]. Consequently, 63% of the annotated units contain no more than 2 prosodic words. These results are in conformity with the MaxBin constraint [10] on the size of phonological phrases, introduced within the paradigm of Optimality Theory. Hence, more than 60% of the perceived units in our study are of the size of phonological phrases. No underlying principle of prosodic phonology forbids that an intonational unit should contain only one phonological phrase, though the application of such a strategy induces a stylistic effect: the listeners perceive a highly solemn, elevated style [1]. The style of the analysed speech

data is far from that of rhetorical speech. Therefore, we can suppose that an intermediate level of phrasing is relevant for the prosodic analysis of Russian, namely the level of phonological phrases, which receive a special acoustic marking in speech signal.

Table 3. Proportions of units of different length

Language	Condition	3 syllables and less	6 syllables and less
Russian	CondF	12.632	71.754
	CondD	9.160	60.878
	CondN	13.910	62.782
French	CondF	10.448	62.313
	CondD	9.542	63.931
	CondN	10.448	64.225



**Figure 3:** Prosodic boundary after “yes” particle without support from melodic cues

### 3.3. Temporal organisation of perceived units

Next, we analysed the data on the temporal organisation of perceived units. In our data, the mean unit duration exhibits small variations both between experimental conditions and in the performance of two groups of participants; it is of 0.86 s. Note that in the Russian prosodic literature the authors communicate the mean durations of syntagmas as of 1.2 s. The units annotated in our study are consequently smaller than “classic” syntagms, this fact indicating once again that intermediate units have been annotated (phonological phrases).

One more hypothesis was tested as to the duration of the units. We noted that there is a difference in the temporal structure between non-terminal and terminal units: several listeners marked shorter units at the beginning of the stimuli. The hypothesis of the difference in duration between non-terminal and terminal units was tested statistically with mixed linear models. As several hypotheses were tested simultaneously, we applied a Bonferroni correction to the relevant p-values: hence, in this analysis, a contrast is significant if  $p < 0.017$  ( $0.05 / 3$ ).

As shown by the data summarised in table 4, in the performance of Russian speaking subjects the non-terminal units are significantly shorter than terminal ones under CondF and CondN. For the French speaking subjects the durational differences are significant under all three conditions.

Table 4. *Differences between non-terminal and terminal units*

Language	Condition	$\Delta_{T-NT}$ (ms)	t value	p value
Russian	CondF	307.7	8.028	0.0000
	CondD	35.33	0.904	0.366
	CondN	143.4	3.688	0.0002
French	CondF	329.35	9.445	0.0000
	CondD	221.65	6.298	0.0000
	CondN	191.95	5.532	0.0000

These data agree with the procrastination constraint related to the structuring of speech in production. Introduced by Wasow [12], this constraint stipulates that the production of complex/longer units is postponed as more planning effort is required. Our data, obtained in a perception study with delexicalised stimuli agree with this hypothesis.

#### 4. Discussion and Conclusions

The goal of our perception study was to test a number of hypotheses about prosodic constituency in Russian. In the analysis of the results we took into account both phonological constraints on prosodic phrasing and details of phonetic (temporal) organisation of the perceived units. We presented the evidence in favour of distinguishing two intermediate levels in the prosodic constituency of Russian inferior to Intonational Units: metrical domains (accentual phrases) and phonological phrases. We cannot advance any hypothesis about the status of metrical domain, namely, whether it is co-extensive with a prosodic word or rather with a metrical foot: note that in the Russian prosodic research the references to the concept of foot are rare. This issue should be clarified in further studies.

A great part of annotated units seem to be phonological phrases. Consequently, we can advance the hypothesis that phonological phrases receive a special acoustic marking and probably are supported by pitch cues. In future work it is planned to answer the question of how intonation boundaries are realised at different levels of the prosodic hierarchy.

The main conclusions of our work are largely motivated by the eurhythmic constraints on the formation of prosodic constituents. As we pointed out, these constraints are to be taken into account together with syntactically and pragmatically oriented constraints. At the same time, the inter-subject variations in the phrasing decisions reveal that listeners exhibit different strategies in parsing the speech ("metrical" versus "intonational" strategies were identified). To account for the phrasing strategies attested in Russian further studies are required, which will allow the modelling of complex interfaces.

#### 5. References

[1] Nespor, M. and Vogel, I., *Prosodic Phonology*, Foris Publication, Dordrecht, 1986.  
 [2] Beckman, M.E. and Pierrehumbert, J.B., "Intonational Structure in Japanese and English", *Phonology Yearbook*, 3:255-309, 1986.  
 [3] Sanderman, A.A. and Collier, R., "Prosodic rules for the implementation of phrase boundaries in synthetic speech", *J. Acoust. Soc. Amer.*, Vol. 100(5), 1996, 3390-3397.

[4] Frazier, L., Clifton, Ch. Jr., and Carlson, K., "Don't break, or do: prosodic boundary preferences", *Lingua*, Vol. 114(1), 2004, 3-27.  
 [5] Ščerba, L.V., *Izbrannyje raboty po russkomu jazyku*, Moscow, 1957.  
 [6] Selkirk, E., *Phonology and Syntax: The Relation Between Sound and Structure*, Coll. Current Studies in Linguistics Series, 10. Cambridge, MA, USA: The MIT Press.  
 [7] Cohen, J., "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Vol. 20, 1960, 37-46.  
 [8] Gee, J. and Grosjean, F., "Performance structures: A psycholinguistic and linguistic appraisal", *Cognitive Psychology*, Vol. 15, 1983, 411-458.  
 [9] Prieto, P., "Syntactic and eurhythmic constraints on phrasing decisions in Catalan", *Studia Linguistica* (special issue on 'Boundaries in Intonational Phonology'), Vol. 59 (2-3), 2005, 194-222.  
 [10] Selkirk, E., "The Interaction of Constraints on Prosodic Phrasing". In Horne, M. (ed.), *Prosody: Theory and Experiment*, Coll. Text, Speech and Language Technology, 14. Dordrecht, Pays-Bas: Kluwer Academic Publishers, 2000, 231-261.  
 [11] Stern, A. S., "Some statistical characteristics of Russian spontaneous speech". In Svetozarova, N.D. (ed.), *Phonetics of spontaneous speech*, Saint-Petersburg, Russia, 1988, 196-224 (in Russian).  
 [12] Arnold, J., Wasow, T., Losongco, A., and Ginstrom, R., "Heaviness vs. Newness: The Effects of Complexity and Information Structure on Constituent Ordering", *Language*, Vol. 76, 2000, 28-55.