РУССКИЙ ИНТОНАЦИОННЫЙ КОРПУС: ПРЕДВАРИТЕЛЬНЫЙ ОТЧЁТ[1]

Архипов А. В. (sarkipo@yandex.ru),
Захаров Л. М. (leon@philol.msu.ru),
Кривнова О. Ф. (okri@philol.msu.ru),
Кодзасов С. В. (sankod@yandex.ru),
Лебедев А. А. (andrettile@inbox.ru)

Московский государственный университет имени
М. В. Ломоносова, Москва, Россия

Ключевые слова: мультимедийный корпус, интонация, просодия, русский язык

RUSSIAN INTONATION CORPUS: A PRELIMINARY REPORT

Arkhipov A. V. (sarkipo@yandex.ru),
Zakharov L. M. (leon@philol.msu.ru),
Krivnova O. F. (okri@philol.msu.ru),
Kodzasov S. V. (sankod@yandex.ru),
Lebedev A. A. (andrettile@inbox.ru)

M. V. Lomonosov Moscow State University, Moscow, Russia

The paper presents Russian Intonation Corpus (RINCO), the multimedia speech corpus designed for the study of prosody in Russian, its history, purpose, and functionality. RINCO started as a series of MS Access databases developed in 2004–2009 by a research team at the philological faculty of Moscow State University led by S. V. Kodzasov. The databases contained descriptions of most important prosodic characteristics (several tonal features, tempo, loudness, phonation types) for quasi-natural dialog utterances, as well as for a selection of spontaneous and prepared spoken narrative texts of different genres. These descriptions have been transferred to a multi-layer time-aligned markup format, suitable for local use in ELAN annotator and for online browsing, playback and search using the LAT server platform. The corpus in its current state should be fully operational at http://languedoc.philol.msu.ru/rinco by the end of 2012. Various possibilities of enhancing the corpus are discussed, as well as some findings with a sample query for "IK-5" intonational construction.
The project is currently supported by RFH grant #11-04-12042.

Keywords: multimedia corpus, intonation, prosody, Russian language

---

1.   Project overview

The Russian Intonation Corpus (RINCO) originated in 2004 as a project for creating an array of dialogical utterances provided with complex prosodic markup. The scale of the data and the nature of the markup were adapted for the purposes of linguistic investigations rather than direct use in speech technologies. In the following three years (2004–2006), the team led by Sandro V. Kodzasov developed three MS Access databases, each devoted to a specific type of utterances (questions, imperatives, statements). Another three-year project was undertaken to analyze monological spoken texts of different genres or their fragments: TV news programs, literary and traditional prose reading, and real life stories (2007–2009). These two projects were supported by RFH grants # 04-04-12027v and # 07-04-12160v and presented in a number of previous publications in DIALOG proceedings (see e. g. Kodzasov et al. 2005, 2006, 2008).

The goal of the current project is to bring the previously collected data to a new level of efficiency and accessibility. The main issues addressed are (1) open web access and feedback, (2) time-alignment of the prosodic features' markup, and (3) improved search capabilities. In order to attain (1)–(3), the MS Access databases were ported to a new platform, consisting of ELAN (for local use) and the Language Archiving Technologies (LAT) software suite (for online use), both developed at Max-Planck Institute for Psycholinguistics, Nijmegen (Netherlands). At the same time, we tried to enhance the markup with some new tiers of information, as well as to develop documentation for a wider audience (in progress).

The current project, scheduled to finish in 2012, will lay ground for a full-scale multimedia speech corpus of Russian for intonation studies which can be further enhanced in several directions, including dataset expansion, introducing new kinds of markup (e. g. alternative transcription systems), improving metadata, navigation and help resources, user interface and export options.

2.   Corpus presentation

2.1. Main features

The main features that make RINCO special compared to other existing corpora are the following:
- RINCO is centered around the recordings of speech, not around its transcripts (unlike e. g. the spoken subcorpus of the Russian National Corpus).
- The main motivation behind RINCO was to illustrate the prosodic variation in spoken Russian and thus to cover the most diverse utterance types, rather than to construct a weighted representative corpus.
- For the same reason, the data for the first (dialogical) subcorpus were recorded in studio rather than taken from spontaneous speech recordings, since observation of natural speech would necessitate huge volumes of data to cover all the desired utterance types.
- The prosodic markup used in RINCO follows the combinatory model of prosodic features by Sandro V. Kodzasov [Kodzasov 1996].

- The target audience includes primarily researchers in phonetics & phonology as well as in semantics & pragmatics. RINCO was not designed specifically for use in ASR or TTS systems; however, we hope that fruitful partnership is possible with experts in speech technologies. Also, RINCO can also be useful for students and teachers of Russian (both as second and first language, e. g. as an aid in mastering intonation for people with hearing impairments).

2.2. Data

The first subcorpus, "Intonation of Russian Dialogue" (IRD), includes around 1,000 utterances of three major types (questions, imperatives, and statements), with around 300–400 utterances for each type, totalling ca. 90 minutes recording time. These are utterances pronounced in isolation or in "cue-response" pairs. The textual examples that were recorded originate from authors' observations and recordings or published linguistic works, partially adapted or constructed to cover most important semantic, pragmatic, and grammatical types. The speakers who participated in recording sessions were Moscow inhabitants with higher education, all native speakers of Standard Russian, both men and women, aged from 20 to 70.

The second subcorpus included (fragments of) monological discourse of different genres: live TV news programs, literary and traditional narrative texts (novels, folk tales) recorded by professional readers/actors, and real life stories told by several speakers. The overall time of the monological subcorpus is around 90–100 minutes.

The primary data was principally recorded in Linear PCM .wav format (except for the database on questions which was first recorded on magnetic tape and then digitized).

2.3. Markup

The linguistic annotations provided for the speech samples include multiple layers (tiers) of markup, time-aligned to the speech signal where appropriate. The tiers used in the IRD subcorpus include:
- *Text* — Text of the utterance in orthographic representation.
- *Transcribe* — Canonical phonetic transcription as generated by the transcriber component of the NRS speech synthesis package. This is not (yet) time-aligned or manually checked.
- *Voice* — Male or female.
- *Int phrase* — Same as Text, segmented into intonation phrases.
- *Words* — Same as Text, segmented into words.
- *Accent* — Local tonal accents (includes localization and directions of pitch movement, interval of pitch change and some non-tonal prosodic accents (e. g. loudness or duration)).
- *Tone* — All non-automatic pitch movements.
- *Pause* — Pauses longer than 0.2 sec
- *Loudness, Tempo, Register, Phonation* — Integral prosodies.

- *Type 1, Type 2* — Utterance type according to the complex semantic and grammatical classification developed by project members. Only the highest and the lowest levels of classification are included in the annotation file itself; however, all the levels of classification are represented in (browsable and searchable) metadata which form the tree structure of the corpus (see below).
- *Comments*
- *ID* — Internal code of the utterance.

The monological subcorpus uses the same tiers except for the classification tiers (Type 1, Type 2).

2.4. Setup for local use

The markup was implemented in ELAN multimedia annotator [Wittenburg et al. 2006]. ELAN was initially designed for the purposes of studying sign language and is also widely used in various endangered language documentation projects. ELAN supports an arbitrary number of text annotation tiers time-aligned to the sound (and/or up to four video files). ELAN can also display pitch graphs generated by the Praat phonetic analyzer (*.PitchTier). Annotation tiers can be organized in tree-like hierarchy with several possible types of relation between them.
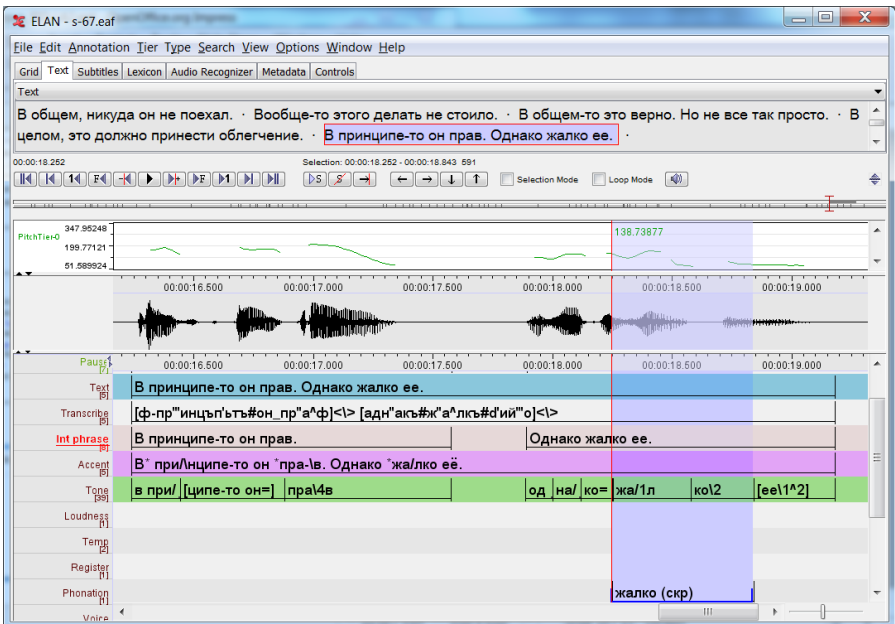

Fig. 1. Sample ELAN window

ELAN can be used to create, edit, and browse the annotated media files, as well as to run complex search queries within a single file and across large collections of files. The search engine allows for searching using regular expressions, searching by specifying various constraints on structural and temporal relations between annotations within the same tier or across multiple tiers. Search results can then be explored in ELAN or exported e. g. as CSV tables.

2.5. Online implementation

The online version of the corpus uses the Language Archiving Technologies (LAT) platform which includes several pieces of software, as the IMDI browser, serving as the main interface to the corpus, and the ANNEX utility used for online access to the annotation files, offering much of the ELAN functionality to explore (not edit) ELAN files (and some other kinds of text annotation formats) online, except for online display of pitch graphs. In particular, ANNEX allows for playing back the media fragment corresponding to the selected time interval or annotation unit. A separate utility, TROVA, offers an interface to the search engine (same as used with ELAN).
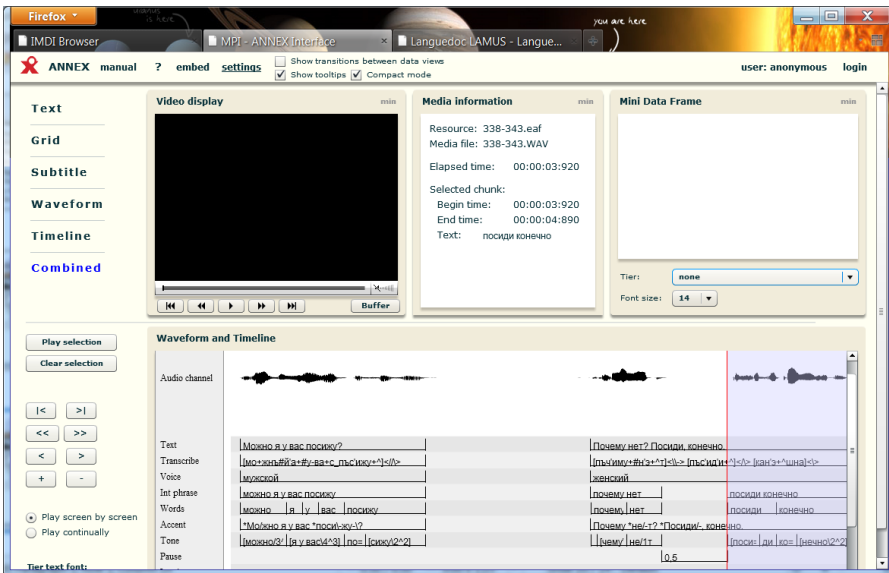

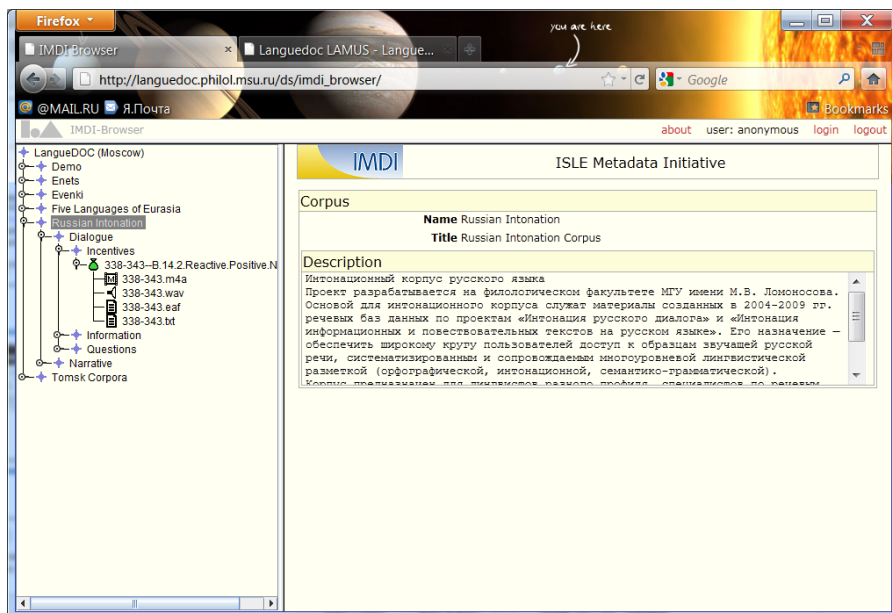**Fig. 2.** Russian Intonation Corpus online: ANNEX viewer

Fig. 3. Russian Intonation Corpus online: IMDI metadata browser

The IMDI browser presents the corpus structure which is calculated based on metadata descriptions of the corpus files. In the case of RINCO, the metadata include information about the project, the speakers/sources of data, media files, the types of utterances (for the IRD subcorpus) according to the adopted classification, the textual content (for the monological subcorpus), and on the access rights. The metadata can also be searched to form a subcorpus (which can be further searched by content).

The IMDI metadata are by principle visible to all users; the corpus creators can further specify access rights to particular files or file types. RINCO offers online access for all users to browse the annotations and play back media files, which means that ELAN annotation files and Praat pitch graphs can be freely downloaded for local use. Access to sound files will be available on demand.

3. An illustration: The uses of "IK-5"

Let us consider a simple search example and look for the tonal "hat contour," traditionally termed "IK-5" ("intonational construction 5") [Bryzgunova 1977: 78; 1980: 116]. A standard IK-5 is usually described as having two nuclei with rising tone on the first one, falling tone on the last one and a level tone (no significant pitch change) between them. This can be searched for as a sequence of "/-" and "-\" tonal accents in the Accents tier, or as a sequence of "/", "=" and "\" movements (in adjacent annotations) in the Tone tier (see Fig. 4).

IK-5 has been initially described as typical for a particular utterance type, namely for exclamatives of admiration, especially with WH-words (hereafter the rise and fall are marked with /- and -\ after the vowel, respectively; the dash — signals the absence of an automatic fall after rising, as opposed to the unmarked / and \ tonal accents):

(1) *Како/-е у-\тро! 'What a (wonderful) morning!' (from [Bryzgunova 1980: 116])*

(2) *За/-меча-\тельно! 'Wonderful!' (from [Bryzgunova 1980: 116])*

However, subsequent studies emphasized that these exclamatives would normally call for a particular phonation type (breathy voice, sometimes also lax voice, and slower tempo) [Kodzasov 1996/2009: 21]. On the other hand, the same tonal configuration with modal voice was shown to be typical of WH-questions [Nikolaeva 1977]. We do indeed find both of these kinds of uses in RINCO, as illustrated in (3) and (4); utterance ID in the corpus is given in square brackets.

(3) *У неё прекра/-сный вку-\с! 'She (has) an excellent taste!' [st79f]*

(4) *Покажите, пожалуйста, ка/-к им по-\льзоваться? '(Could you) please show how to use it?' [i95m]*
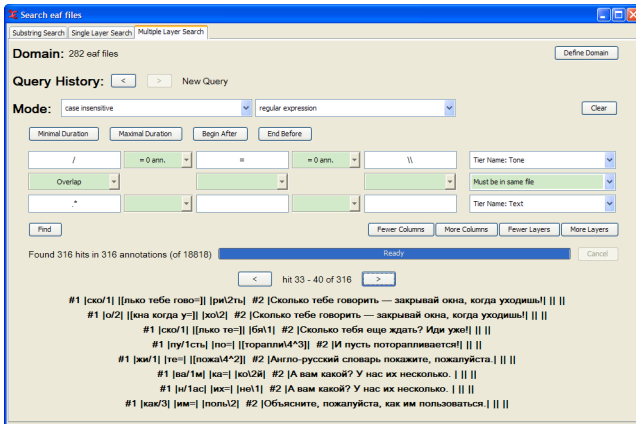

Fig. 4. A sample multi-tier search window

Among questions, one can mention highly emotional questions with astonishment or indignation, which usually demonstrate a greater range of pitch change (above 7–8 semitones):

(5) *Неуже/-ли он об этом не зна-\л? 'Did he really not know about it?' [q122m]*

(6) *Отк/-уда мне зна-\ть? 'How would I know?' [q195f]*

(7) *Заче/-м мне было его бра-\ть? 'Why would I take it?'[q215m]*

(8) *А что/- я такого сказа-\л? 'What did I say?' [q216m]*

However, a significant number of examples falls outside of these classes. We find, for instance, a number of similar configurations in requests, especially in a more formal setting:

(9) *Бу/-дьте любе-\зны, говорите громче! 'Please speak louder!' [i5Af]*

(10) *Прошу/- сади-\ться. 'Please be seated.' [i152m]*

(11) *Англо-русский словарь покажи/-те, пожа-\луйста. '(Could you) please show me the English-Russian dictionary.' [i94Am]*

(12) They can also appear on (cooperative) forms of address, corroborating what had been reported in [Kodzasov 1998/2009: 163]:

(13) *Пё/-тр Ива-\нович, за/-втрака-\ть! 'Petr Ivanovich, (please come) to have breakfast!' [i353f]*

Another use of 'hat contour' is in replies like (13):

(14) *{Полкило творога, пожалуйста.} — А о/-н не очень све-\жий. [i57Af]*
'{Half a kilo cottage cheese, please.} — (But) it is not very fresh.

In general, the combination of rise on Topic and fall on Focus is quite common in statements. However, sustaining high level tone after the rise seems to be correlated with statements used as an indirect negative reply; cf. the following pair:

(15) *Ивано/в заболе\л. 'Ivanov fell ill.'*

(16) *{Иванова можно к телефону?} — А Ивано/-в заболе-\л.*
'{Can I talk to Ivanov(, please)?} — Ivanov has fallen ill.' [Kodzasov 1996/2009: 22]

Apart from the rise-level-fall configuration, one can also observe a comparable number of the opposite contour, fall-level-rise. It is particularly common in questions with Topic preceding Focus, in which case the Topic adapts inversely to the rise on the Focus [Kodzasov 1996/2009: 22]:

(17) *Посу\-ду помо-/ешь? 'Will you wash the dishes?' [i115f]*

(18) *Ча\-йник не поста-/вишь? 'Will you turn on the kettle?' [i117.2m]*

(19) Cf.: *Ты не поставишь чайник? 'Will you turn on the kettle?' [i117.1f]*

(20) *Васю пригласить? 'Shall I invite Vasia?' [i380Af]*

When a question word bearing raising tone precedes the Topic (inversely marked with falling tone), a rise may appear on the utterance-final syllable to keep marking it as a question:

(21) *Что/ до\ма/? 'What (news) at home?' [q143f]*

With this quick glance of IK-5 and related tonal contours, we hope to have demonstrated the usefulness of the possibility to search for specific prosodic patterns, which can serve teachers and students of Russian intonation as well as researchers.

4.   Current state and outlook

The current project is scheduled to finish by the end of 2012, when all the data should have been uploaded to the project server (http://languedoc.philol.msu.ru/rinco). However this will only be the beginning of work on the large-scale corpus of Russian intonation. Several directions of future work can be foreseen.

Most naturally, the range of primary data should be expanded, possibly with recordings of natural (spontaneous) speech of different genres, including natural dialogs.

The prosodic markup can be further enhanced in several ways, namely adapting the automatic phonetic transcription to the real pronunciation; using the phonetic representation as the basis for prosodic markup (instead of the orthographic representation); introducing alternative prosodic transcription schemes — e. g. the ToRI system ([Ode 2003]) and the widely used traditional Russian notation by Elena A. Bryzgunova [Bryzgunova 1977]. The latter will require collaboration by experts in corresponding transcription systems, which can be easily implemented thanks to the customizable markup format of ELAN.

Finally, user experience may be improved by providing detailed metadata descriptions, various navigation aids and help resources, developing the user interface(s) and additional export options.

References

1. *Bryzgunova, E. A.* Sounds and intonation of Russian speech [Zvuki i intonatsiia russkoi rechi]. Moscow, 1977.
2. *Bryzgunova, E. A.* Intonation [Intonatsiia], in Russian Grammar [Russkaia grammatika], v. 1. Moscow, 1980. P. 96–120.

3. *Kodzasov S. V.* Combinatory model of the sentence prosody. [Kombinatornaia model' frazovoi prosodii], in Prosodic structure of Russian speech [Prosodicheskii stroi russkoi rechi]. Moscow, 1996. Reprinted in Kodzasov 2009.

4. *Kodzasov S. V.* The prosody of forms of address. [Prosodiia obrashchenii], in Language: variability and stability [Iazyk: izmenchivost' i postoianstvo]. Moscow, 1998. Reprinted in Kodzasov 2009.

5. *Kodzasov S. V.* Studies in Russian prosody [Issledovaniia v oblasti russkoi prosodii]. Moscow, 2009.

6. *Kodzasov S. V., Bonch-Osmolovskaia A. A., Zakharov L. M., Kobozeva I. M., Krivnova O. F.* Database "Intonation of Russian Dialogue": interrogative utterances [Baza dannykh "Intonatsiia russkogo dialoga": voprositel'nye repliki]. Trudy Mezhdunarodnoi Konferentsii "Dialog 2005" [Proceedings of the International Conference "Dialog 2005"]. Moscow, 2005. P. 245–249.

7. *Kodzasov S. V., Arkhipov A. V., Bonch-Osmolovskaia A. A., Zakharov L. M., Krivnova O. F.* Database "Intonation of Russian Dialogue": imperative utterances [Baza dannykh "Intonatsiia russkogo dialoga": pobuditel'nye repliki]. Komp'iuternaia Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoi Konferentsii "Dialog 2006". [Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2006"] Bekasovo, 2006. Moscow: RSUH, 2006. P. 236–242.

8. *Kodzasov S. V., Arkhipov A. V., Zakharov L. M., Krivnova O. F.* Database "Intonation of Russian Informational Texts" [Baza dannykh "Intonatsiia russkix informatsionnykh tekstov|"]. Komp'iuternaia Lingvistika i Intellektual'nye Tekhnologii: Po materialam ezhegodnoi mezhdunarodnoi konferentsii "Dialog 2008". [Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"] Bekasovo, 2008. Issue 7 (14). Moscow: RSUH, 2008. C. 206–208.

9. *Nikolaeva T. M.* Sentence intonation of Slavic languages [Frazovaia intonatsiia slavianskikh iazikov]. Moscow, 1977.

10. *Odé, C.* Developing a Transcription of Russian Intonation (ToRI). Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, August 3–9, 2003. P. 3097–3100.

11. *Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H.* ELAN: a Professional Framework for Multimodality Research. Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation. 2006. [http://www.lat-mpi.eu/tools/elan/]