

## LITHUANIAN SPOKEN CORPORA AND STUDIES OF FIRST LANGUAGE ACQUISITION: A VIEW FROM OUTSIDE

**Elena Riekhakaynen**

Saint-Petersburg State University  
Universitetskaya emb., 11  
199034 St. Petersburg, Russia  
e-mail: [e.riehakajnen@spbu.ru](mailto:e.riehakajnen@spbu.ru)

### **Introduction**

The idea for this paper appeared within our current research of phonetic reduction in spontaneous speech of Russian monolinguals aged three to five years. We assume that the results of this study will contribute to the understanding of whether reduced word forms are stored in the mental lexicon of a native speaker, how they can enter the mental lexicon during first language acquisition, and how they are processed during speech production and spoken word recognition (see Ernestus et al. (2002), Nigmatulina et al. (2016) for further discussion of the problem).

The first step in any spontaneous speech research is to choose or create a corpus of spontaneous speech. We have been developing a corpus of spontaneous Russian since 2009. At the moment, it includes 115 minutes of radio interviews and television talk shows, along with their orthographic and acoustic-phonetic annotations. The corpus is called the Corpus of Transcribed Russian Oral Texts (CTROT), and the principles of its annotation as well as the annotation itself are available online. We decided to use the same principles of annotation we used in the Corpus of Transcribed Russian Oral Texts for the corpus of child speech. However, the method of obtaining the data for the latter seemed to be problematic. Thus, we have analyzed the experience of creating spoken corpora (in both adults and children) and studying first language acquisition in different languages.

There is a relatively long tradition of studying first language acquisition in Lithuanian; numerous studies have appeared in the field during the last 25 years. For this reason, we decided to find out how studies in Lithuanian can contribute to the methodology of collecting data for first language acquisition studies in other languages. We will analyze several Lithuanian spoken corpora, overview papers on first language acquisition of Lithuanian, and compare them to similar studies of Russian spontaneous speech and first language acquisition.

## **Developing a Corpus of Spontaneous Speech: From Adults to Children or Vice Versa?**

Lithuanian is not usually mentioned when discussing linguistic corpora in general and spoken corpora in particular. Both in the Catalogue of the European Language Resource Association (ELRA) and the Catalog of Linguistic Data Consortium (LDC), there is only one language resource for Lithuanian (ECI Multilingual Text that includes written texts in Lithuanian of the total size of 20 000 lexical words). According to the press release published by the META-NET Project in 2012, Lithuanian was among the languages with weak or no support in the domain of speech and text resources (Vaišnienė, Zabarskaitė 2012, 75). However, in the same book (p. 62), one can find a brief overview of Lithuanian spoken corpora, showing that spoken corpora have been developing in Lithuania since the beginning of the 1980s. Moreover, some new resources that will be described below have appeared after 2012.

Although they are not numerous, spoken corpora of Lithuanian have quite a strong methodological advantage — their structure and annotation correspond well to their intended theoretical or practical usage. For example, Laurinčiukaitė et al. (2018), while working on the problem of automatic speech recognition and synthesis, created the corpus *Liepa* consisting of two parts (the corpus can be requested free of charge by filling in the form at the site of the project LIEPA). Its first part is aimed to be used in speech recognition and comprises records of 376 native speakers of Lithuanian reading out words, phrases and texts. The amount of speech in the corpus is 100 hours. The developers of the corpus claim that “the main efforts should be given to achievement of high accuracy of speech recognition” and thus include in this part of the corpus only “speech data without significant noise, phonetic distortions of words” (Laurinčiukaitė et al. 2018, 491). For the second part, aimed at being used for speech synthesis, 13 hours of records of even higher quality were collected. These are sentences including all features of “Lithuanian acoustic space” pronounced by four speakers (two men and two women). The problem of the quality of speech is one of the most crucial for the development of a spoken corpus. The above discussed paper is an example of one of the possible ways to solve this problem in accordance with the main goal of the project - to provide the data for automatic speech recognition and synthesis. However, working only with the data of high quality, we risk missing or underestimating some problems that occur in natural speech (such as reduction, contractions at word boundaries, etc.). Consequently, such an automatic model will be able to recognize only prepared clear speech. For this reason, large corpora of spontaneous speech are also required.

At least two open-access spoken corpora of Lithuanian include spontaneous speech. They are the Database of Spoken Language (Sakytinės kalbos įrašų bazė; DSL) and The Corpus of Spoken Lithuanian (Sakytinės lietuvių kalbos tekstynas; CSL). The information about the former is mainly in Lithuanian (Kazlauskienė, Raškinis 2013). By 2012, this corpus comprised – inter alia – ten hours (14 100 words) of spontaneous speech from ten men and twelve women. It is available for online-search. There are options to search for words as well as for noise, laugh and certain hesitations. The results are presented in orthography, in KWIC format. For some words, there is an option to see their pronunciation variants. There are also icons for listening to a selected word string or a separate word, but the sound is not available for some reason. The moment this technical problem has been solved, this corpus will become an example of a good spoken corpus, as it allows its users both to read the orthographic annotation of a word string and to listen to the sound. Users are not only able to listen to the whole utterance, but also to selected isolated words. To compare: the option to listen to a separate word from an utterance is still not realized in any of the Russian spoken corpora available online (although in MURCO there is an option to download the sound of the whole utterance). The Phonetic Corpus of Estonian Spontaneous Speech (PhCESS), which has several levels of annotation (including phonetic transcription) and the option to listen to the sound, allows users to download the sound files and annotations (TextGrid files) only for the whole clauses (sequences of words), not separate words or sounds.

What makes The Corpus of Spoken Lithuanian (Sakytinės lietuvių kalbos tekstynas) special is the assumption that “systematic research of spoken Lithuanian is closely related to the development of child language corpora” (Dabašinskienė, Kamandulytė 2009, 67). The Corpus of Spoken Lithuanian contains around 100 hours of records (300 000 words), 60 of them being spontaneous (private or institutional) and 40 – prepared (Kamandulytė-Merfeldienė 2017, 875). It is a rare example of the situation when the developers of an adult speech corpus have taken into consideration the principles of annotation used in a corpus of child speech (the Lithuanian Child Language Corpus, which also includes child-directed speech, so we will further refer to it as the Lithuanian Corpus of Child and Child-Directed Speech). For the Russian language we have only one such example (CNDS; Кибрик, Подлесская 2009) where all the texts were annotated using the principles developed for the Corpus of Night Dream Stories of children aged seven to seventeen years.

The development of the Lithuanian Corpus of Child and Child-Directed Speech started in 1993 within the international project “Crosslinguistic Project on Pre- and Protomorphology in Language Acquisition”. According to Dabašinskienė, Kamandulytė (2009, 69), the corpus includes about 200 hours of longitudinal recordings from four children and their

caregivers. The records started when the children were 17-19 months old and lasted for about a year (one boy started later - being 25 months old, and the records lasted for more than two years). The conversations were recorded by the parents in different settings.

The Lithuanian Corpus of Child and Child-Directed Speech as well as the Corpus of Spoken Lithuanian are annotated according to the principles proposed in the Child Language Data Exchange System (CHILDES). The rules of annotation and some methodological problems are discussed in Dabašinskienė, Kamandulytė (2009); Kamandulytė-Merfeldienė, Balčiūnienė (2016); Kamandulytė-Merfeldienė (2017). As the borders between utterances are not always clear in spontaneous adult speech, there is a theoretical and methodological problem of a unit of annotation (transcription unit). The same problem for the Russian language is discussed in Раева, Риехакайнен (2015); Nigmatulina et al. (2016) where the authors propose to use intervals between pauses rather than utterances or clauses while transcribing spontaneous speech. Although the utterances are generally much shorter in child speech than in adult speech, this problem also should be taken into account while creating a corpus of child speech (especially for elder children who are already able to produce quite long narratives - see, for example, Eismont et al. (2018)).

Besides the problem of an annotation unit, the following methodological solutions, found while developing the spoken corpora of Lithuanian, seem to be relevant for creating a corpus of spontaneous speech for other languages, and especially for our study of reduction in children speech:

- the decision to collect more data than is actually needed and to record conversations that are at least 15-30 minutes long allows to find both a certain number of grammatical constructions under research (and that was the “task” of the developers of the Lithuanian Corpus of Child and Child-Directed Speech) and different realizations of one and the same word (that is in focus of our current research);
- the principle of naturality: due to the aim to collect the most natural speech, the developers of the Corpus of Spoken Lithuanian told their informants that their speech had been recorded only after the recording session; both children and adults (for the spontaneous part of the corpus) were recorded in natural communication situations; although the parents in our research do sign a written form of consent for their children to take part in the research, we also ask parents to organize recording sessions so that the children do not know that their speech is being recorded;
- the idea to transcribe child speech both orthographically and phonetically (whereas the speech of their adult communicants is transcribed only orthographically).

Unfortunately, there is no Lithuanian data in the CHILDES Browsable Database. So, as far as we understand, the Lithuanian Corpus of Child and Child-Directed Speech is now available only for the researchers who were involved in the project.

The Corpus of Spoken Lithuanian (*Sakytinės lietuvių kalbos tekstynas*) is available online. The search options include lemma and word-form search, choosing the type of spoken discourse (private and institutional spontaneous speech being in the list), the place of the recording, and personal characteristics of speakers (male / female and age). It is worth mentioning that the age of the informants whose speech is included in the corpus is from three to 81 years and the youngest age group is from zero to 11 years. It means that the Corpus of Spoken Lithuanian includes not only adult, but also child and adolescent speech. According to the search results, there are more than 19 000 words in the subcorpus of spontaneous speech of children aged between three and 11 years and around 9 500 words in the subcorpus of spontaneous speech of adolescents between 12 and 18 years. Thus, the Corpus of Spoken Lithuanian has potential to be used for the developmental studies of spontaneous speech processing. The idea to include child and adolescent speech into the spoken corpus of a certain language along with the recordings of adult speech is promising, especially if the principles of annotation for all types of speech are the same. From this point of view, the methodological assumptions of our project are close to those of the Corpus of Spoken Lithuanian. We also annotate (both orthographically and phonetically) the Russian child speech the same way we annotated the spontaneous adult speech (Nigmatulina et al. 2016) for further details) and are going to include these data into the Corpus of Transcribed Russian Oral Texts.

Discussing Lithuanian spoken data from different age groups, we should also mention the SACODEYL corpus, which comprises video interviews of 13-17-year-old native speakers of seven European languages, Lithuanian being one of them. The corpus of each of these languages includes 20 to 25 interviews of approximately ten minutes each. All videos have orthographical transcripts as well as some pedagogically oriented annotation (topic, discourse markers, etc.) because the main goal of the corpus is to provide teachers and learners with the material for learning a language (Widmann et al. 2011). Unfortunately, the search option on the web page of the project does not work for the moment. But as soon as it becomes available it can be used not only by teachers and students, but also by researchers interested in comparative studies of spoken language of different age groups.

### **First Language Acquisition: Evidence from Lithuanian**

The data collected for the Lithuanian Corpus of Child and Child-Directed Speech has been studied in a number of papers. Many of them are about the acquisition of grammar and

lexico-grammatical or lexical groups. The researchers pay special attention to the functional aspect trying to figure out how the acquisition of a certain grammatical category or lexical group contributes to language acquisition in general. For example, I. Savickienė argues that the acquisition of diminutives helps Lithuanian children to acquire declensional noun endings (Savickienė 2007). L. Kamandulytė-Merfeldienė shows that “children start using diverse forms of adjectives in multiword utterances only when they acquire agreeing features and the structure of agreeing combinations (attributive and predicative)” (Kamandulytė-Merfeldienė 2013, 99).

As the Lithuanian Corpus of Child and Child-Directed Speech was developed within a crosslinguistic project, there are several papers where Lithuanian data is compared to the Russian longitudinal material (Balčiūnienė, Kazakovskaja 2011; Voeikova, Dabašinskienė 2012; Казаковская и др. 2013; Dabašinskienė, Voeikova 2015, etc.). The examples of studies including several languages (Lithuanian being among them) are the following: Savickienė, Dressler (2007); Tribushinina et al. (2013). Such crosslinguistic projects allow the authors to put forward hypotheses about universal and language specific tendencies in first language acquisition. Lithuanian and Russian, being structurally different from, for example, English (that is more often in focus of first language acquisition studies), can, of course, provide new evidence on how children acquire different aspects of speech and language.

Comparative studies based on the Lithuanian Corpus of Child and Child-Directed Speech data are not restricted to the description of child language: child-directed speech is analyzed as well, and such research shows that the frequency of certain phenomena in the speech of caregivers influences the realization of these phenomena in the speech of a child (Kazakovskaya, Balčiūnienė 2012; Казаковская и др. 2013; Казаковская, Балчюнене 2016, etc.).

From a methodological point of view, it is worth saying that many of the above-mentioned studies do not make use of the whole amount of longitudinal data from the Lithuanian Corpus of Child and Child-Directed Speech. Most often, only the speech of one child is analyzed (in comparative research - in parallel with the data of one child speaking another language). A rare exception is Kamandulytė-Merfeldienė's (2013) study, where adjectives in the speech of all the four children are described.

Although child speech in this corpus is not only orthographically, but also phonetically annotated, the phonetic and/or phonological studies of the data are sporadic. In Kamandulytė (2006, 88), the author says that “the phonetic, phonological, syntactic or lexical features of Lithuanian first language acquisition have not yet been investigated”. Since

that time several papers have appeared that can be regarded as at least partly phonetic or phonological. For instance, L. Kamandulytė (2006) focused on morphotactics and showed that consonant clusters are acquired easier by a child if there is a morpheme boundary inside of them. In Kamandulytė-Merfeldienė (2015), the author develops the idea using experimental rather than corpus data from children aged three to seven years (in this study, the Lithuanian Corpus of Child and Child-Directed Speech is used as a source of examples of words with consonant clusters to be tested in the experiment). E. Krivickaitė (2016, 5) argues that phonetic research in the domain of Lithuanian as a first language are still not frequent and also proposes an *experimental* study of children between four and almost nine years old in order to study the acquisition of phonotactics. The experimental paradigms used by the above-mentioned authors include production tasks (with pictures and stimulus sentences) and word or non-word repetition tasks. This shift towards experimental methods for studying the speech and language acquisition of elder children is not specific to the phonetic domain. I. Balčiūnienė, for example, also prefers experiments (story-telling response tasks) in her numerous papers on the narrative development of pre-school and school-age children (Balčiūnienė 2011; Balčiūnienė 2013, etc.). It seems that such a combination of corpus and experimental data can finally lead to a more detailed description and understanding of the acquisition of Lithuanian as a first language. The data on the acquisition of phonology by Lithuanian children also include some observational examples from how children process language while playing language games. Such examples normally serve as supportive evidence for the phonological assumptions of the authors (see Girdenis (2014, 133) for Lithuanian; Маркус, Гирденис (2011) for Latvian and Lithuanian; Касевич (1981, 144) for Russian, etc.).

Unfortunately, there is no Lithuanian data in the Wordbank, which compiles the data from MacArthur-Bates Communicative Development Inventory (MB-CDI) - parent-report questionnaires for 29 languages. So, this method of studying first language both within one language and cross-linguistically is not available for Lithuanian for the moment.

### **Concluding Remarks and Perspectives**

In this paper, we overviewed the Lithuanian spoken resources and studies of first language acquisition that can be used as methodological inspiration for researchers working with other languages.

The idea of developing the corpora of adult and child speech using the same principles is methodologically appealing, as it will allow researchers to compare spontaneous speech of different age groups. We also have mentioned some technical hints while analyzing the

Lithuanian spoken resources, such as an option to listen online to a single word within a given utterance from a spoken corpus. The majority of the questions discussed by Lithuanian corpus linguists are similar to the problems that appear while developing spoken corpora of other languages. For example, the problem of the boundaries of an utterance is crucial for Russian spoken corpora as well.

The information about the availability of the Lithuanian digital resources provided in this paper may be useful for scholars who do their research on Lithuanian and for students who study Lithuanian as a second language. It seems that since the report of META-NET Project in 2012, the Lithuanian language has become more influential in the domain of speech and text resources. However, the research of the first language acquisition (not only for Lithuanian, but in general) would benefit from the Lithuanian data in the MB-CDI and the Browsable Database of the CHILDES.

The studies of the acquisition of Lithuanian as a first language do use almost all methods that are common in the field: longitudinal studies of the speech of younger children and experiments with elder children (preschoolers and school-age children). In many crosslinguistic studies, Lithuanian is compared to Russian and common tendencies are revealed. As the data from the Lithuanian Corpus of Child and Child-Directed Speech is annotated both orthographically and phonetically, it can (and probably should) be used – together with experimental data from elder children – in phonetic studies that are still not numerous for Lithuanian.

### **Acknowledgements**

The research is supported by the grant number MK-6776.2018.6 from the President of the Russian Federation.

### **Data Sources**

CHILDES - CHILDES Browsable Database, 2018. <https://childes.talkbank.org/browser/>. Accessed 26 December 2018.

CNDS - Рассказы о сновидениях и другие корпуса звучащей речи (the Corpus of Night Dream Stories and Other Corpora of Spoken Speech), 2018. <http://spokencorpora.ru/>. Accessed 26 December 2018.

CSL - Sakytinės lietuvių kalbos tekstynas (The Corpus of Spoken Lithuanian), 2018. <http://sakytinistekstynas.vdu.lt/index.php>. Accessed 27 December 2018.

CTROT - Корпус транскрибированных русских устных текстов (Corpus of Transcribed Russian Oral Texts), 2018. <http://narusco.ru/search/trn-search.php>. Accessed 26 December 2018.

DSL - Sakytinės kalbos įrašų bazė (the Database of Spoken Language), 2018. <http://donelaitis.vdu.lt/garsynas/>. Accessed 25 December 2018.



ELRA - European Language Resource Association, 2018. <http://catalog.elra.info/en-us/repository/search/?q=Lithuanian>. Accessed 25 December 2018.

LDC - Catalog of Linguistic Data Consortium, 1992-2018. <https://catalog.ldc.upenn.edu/search>. Accessed 26 December 2018.

LIEPA - The project LIEPA – Services controlled by the Lithuanian Speech, 2015-2018. <https://www.xn--ratija-ckb.lt/liepa/about-project-liepa/7596>. Accessed 27 December 2018.

MURCO - Мультимедийный подкорпус Национального корпуса русского языка, 2003-2018. <http://ruscorpora.ru/search-murco.html>. Accessed 26 December 2018.

PhCESS - Phonetic Corpus of Estonian Spontaneous Speech. <http://www.murre.ut.ee/otsing/ekskfk.php>. Accessed 25 December 2018.

SACODEYL - SACODEYL corpus, 2018. <https://www.um.es/sacodeyl/>. Accessed 25 December 2018.

Wordbank - Wordbank: An Open Database of Children's Vocabulary Development, 2018. <http://wordbank.stanford.edu/>. Accessed 25 December 2018.

## References

Balčiūnienė, Ingrida 2013, Fictional Narrative Development: Insights from the Cross-sectional Study of Lithuanian Preschool- and School-age Children, *Проблемы онтолингвистики - 2013: Материалы международной научной конференции 26-29 июня 2013 г., Санкт-Петербург*, Санкт-Петербург: РГПУ им. А. И. Герцена, 268-269, <http://www.ontolingva.ru/ontolingv2013.pdf>.

Balčiūnienė, Ingrida 2011, Narrative skills among Lithuanian preschoolers, *Онтолингвистика - наука XXI века. Материалы международной конференции, посвященной 20-летию кафедры детской речи РГПУ им. А. И. Герцена. 4-6 мая 2011 г., г. Санкт-Петербург*, 279-282, <http://www.ontolingva.ru/ontolingvistika2011.pdf>.

Balčiūnienė, Ingrida, Kazakovskaja Viktorija V. 2011, Klausimų forma ir funkcijos lietuvių ir rusų vaikiškojoje kalboje, *Kalbotyra*, 63 (3), 9-25, <http://www.zurnalai.vu.lt/kalbotyra/article/view/7649/5526>

Dabašinskienė, Ineta; Laura Kamandulytė 2009, Corpora of Spoken Lithuanian, *Eesti Rakenduslingvistika Ühingu aastaraamat*, 5, 67-77, [http://donelaitis.vdu.lt/publications/Dabasinskiene\\_2009.pdf](http://donelaitis.vdu.lt/publications/Dabasinskiene_2009.pdf).

Dabašinskienė, Ineta; Maria Voeikova D. 2015, Diminutives in spoken Lithuanian and Russian: Pragmatic functions and structural properties, *Contemporary Approaches to Baltic Linguistics*, Peter Arkadiev, Axel Holvoet, Björn Wiemer (eds.), [Trends in Linguistics. Studies and Monographs [TiLSM] 276], Berlin: De Gruyter Mouton, 203-234.

Eismont, Polina; Vladislav Metelyagin, Elena Riekhakaynen 2018, Coherence Understanding Through Cohesion Markers: The Case of Child Spoken Language, *Speech and Computer. 20th International Conference, SPECOM 2018, Leipzig, Germany, September 18-22, 2018, Proceedings*, Alexey Karpov, Oliver Jokisch, Rodmonga Potapova (eds.), [Lecture Notes in Artificial Intelligence 11096], 145-154.

Ernestus, Mirjam; Harald Baayen, Rob Schreuder 2002, The Recognition of Reduced Word Forms, *Brain and Language*, 81 (1-3), 162-173, [http://www.mirjamernestus.nl/Ernestus/public/Ernestus\\_Baayen\\_Brain\\_and\\_Language\\_2002.pdf](http://www.mirjamernestus.nl/Ernestus/public/Ernestus_Baayen_Brain_and_Language_2002.pdf).

Girdenis, Aleksas 2014, *Theoretical Foundations of Lithuanian Phonology*, Second, revised and expanded edition of Theoretical Foundations of Phonology English translation by Steven

Young, Vilnius: Vilnius University, [http://www.esparama.lt/es\\_parama\\_pletra/failai/ESFproduktai/2014\\_\\_Theoretical\\_Foundations\\_of\\_Lithuanian\\_Phonology.pdf](http://www.esparama.lt/es_parama_pletra/failai/ESFproduktai/2014__Theoretical_Foundations_of_Lithuanian_Phonology.pdf).

Kamandulytė, Laura 2006, The Acquisition of Morphonotactics in Lithuanian, *Wiener Linguistische Gazette*, 73, 88-96, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.556.4528&rep=rep1&type=pdf>.

Kamandulytė-Merfeldienė, Laura 2017, Grammatically coded corpus of spoken Lithuanian: methodology and development, *Engineering and Technology International Journal of Cognitive and Language Sciences*, 11 (4), 853-857, [https://eltalpykla.vdu.lt/bitstream/handle/1/35912/ISSN1307-6892\\_2017\\_V\\_11\\_N\\_4.PG\\_853-857.pdf?sequence=1&isAllowed=y](https://eltalpykla.vdu.lt/bitstream/handle/1/35912/ISSN1307-6892_2017_V_11_N_4.PG_853-857.pdf?sequence=1&isAllowed=y).

Kamandulyte-Merfeldiene, Laura 2015, Morphonotactics in L1 acquisition of Lithuanian: TD vs. SLI, *Eesti Rakenduslingvistika Ühingu aastaraamat*, 11, 95-109, <http://arhiiv.rakenduslingvistika.ee/ajakirjad/index.php/aastaraamat/article/view/ERYa11.06/334>

Kamandulytė-Merfeldienė, Laura 2013, The acquisition of syntactic functions of adjectives in Lithuanian, *Проблемы онтолингвистики - 2013: Материалы международной научной конференции 26–29 июня 2013 г., Санкт-Петербург*, Санкт-Петербург: РГПУ им. А. И. Герцена, 96-100, <http://www.ontolingva.ru/ontolingv2013.pdf>.

Kamandulytė Merfeldienė, Laura; Ingrida Balčiūnienė 2016, Syntactically coded Corpus of Spoken Lithuanian: Developmental issues and pilot studies. *Studies About Languages*, 28, 92-101, <http://kalbos.ktu.lt/index.php/KStud/article/view/15131/7916>.

Kazakovskaya, Viktoria V.; Ingrida Balčiūnienė 2012, Lithuanian and Russian child-directed speech: Why do we ask young children so many questions? *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 69-89, <http://arhiiv.rakenduslingvistika.ee/ajakirjad/index.php/aastaraamat/article/view/ERYa8.05/5>

Kazlauskienė, Asta; Gailius Raškinis 2013, Intonuoto garsyno kūrimo principai (Principles of development of the intonational annotated spoken corpus), *Žmogus ir žodis*, 15, 101-110, [https://www.researchgate.net/publication/279956607\\_Intonuoto\\_garsyno\\_kurimo\\_principai\\_Principles\\_of\\_development\\_of\\_the\\_intonational\\_annotated\\_spoken\\_corpus](https://www.researchgate.net/publication/279956607_Intonuoto_garsyno_kurimo_principai_Principles_of_development_of_the_intonational_annotated_spoken_corpus).

Krivickaitė, Egle 2016, *Vaikų kalbos fonotaktikos raida (Phonotactics Development of Children's Language)*, Summary of Doctoral Dissertation, Kaunas: Vytauto Didžiojo universitetas, <https://vb.vdu.lt/object/elaba:15787643/>.

Laurinčiukaitė, Sigita; Laimutis Telksnys; Pijus Kasparaitis; Regina Kliukienė; Vilma Paukštytė 2018, Lithuanian Speech Corpus Liepa for Development of Human-Computer Interfaces Working in Voice Recognition and Synthesis Mode, *Informatika*, 29 (3), 487-498, <https://www.mii.lt/Informatika/pdf/INFO1179.pdf>.

Nigmatulina, Julija; Olga Rajeva; Elena Riechakajnen; Natalija Slepokurova; Anatolij Vencov 2016, How to study spoken word recognition: evidence from Russian, *Slavic Languages in Psycholinguistics: Chances and Challenges for Empirical and Experimental Research*, Tanja Anstatt, Anja Gattnar, Christina Clasmeier (eds.), [Tübinger Beiträge zur Linguistik 554], Tübingen: Narr Francke Attempto Verlag, 175-190.

Savickienė, Ineta 2007, Form and meaning of diminutives in Lithuanian child language, *The Acquisition of Diminutives: A cross-linguistic perspective*, Ineta Savickienė, Wolfgang U. Dressler (eds.), [Language Acquisition and Language Disorders 43], Amsterdam / Philadelphia: John Benjamins Publishing Company, 13-41.

Savickienė, Ineta; Wolfgang U. Dressler 2007, *The Acquisition of Diminutives: A cross-linguistic perspective*, [Language Acquisition and Language Disorders 43], Amsterdam / Philadelphia: John Benjamins Publishing Company.

Tribushinina and Huub van den Bergh, Elena; Marianne Kilani-Schoch; Ayhan Aksu-Koç; Ineta Dabašinskienė; Gordana Hrzica; Katharina Korecky-Kröll; Sabrina Noccetti; Wolfgang Dressler 2013, The role of explicit contrast in adjective acquisition: A crosslinguistic longitudinal study of adjective production in spontaneous child speech and parental input, *First Language*, 33 (6), 594-616, [https://serval.unil.ch/resource/serval:BIB\\_2BEC132AF133.P001/REF.pdf](https://serval.unil.ch/resource/serval:BIB_2BEC132AF133.P001/REF.pdf).

Vaišnienė, Daiva; Jolanta Zabarskaitė 2012, *The Lithuanian language in the Digital Age*, Georg Rehm, Hans Uszkoreit (eds), [White Paper Series], Berlin, Heidelberg: Springer-Verlag, <http://www.meta-net.eu/whitepapers/e-book/lithuanian.pdf>.

Voeikova, Maria D.; Ineta Dabašinskienė 2012, What can child language tell about language development? A case for case in Lithuanian and Russian, *Multiple Perspectives in Linguistic Research on Baltic Languages*, Aurelija Usonienė, Nicole Nau, Ineta Dabašinskienė (eds.), Cambridge: Cambridge Scholars Publishing, 43-75.

Widmann, Johannes; Kurt Kohn; Ramon Ziai 2011, The SACODEYL search tool - exploiting corpora for language learning purposes, *New Trends in Teaching and Language Corpora*, Ana Frankenberg-Garcia, Lynne Flowerdew, Guy Aston (eds.), London: Continuum, 321-327, [https://www.researchgate.net/publication/272743598\\_The\\_SACODEYL\\_search\\_tool\\_-\\_exploiting\\_corpora\\_for\\_language\\_learning\\_purposes](https://www.researchgate.net/publication/272743598_The_SACODEYL_search_tool_-_exploiting_corpora_for_language_learning_purposes).

Казаковская, Виктория Виладиевна; Ингрида Балчюнене 2016, Вопросительные повторы реплик в русском и литовском инпуте (к проблеме избыточности), *Acta Linguistica Petropolitana. Труды института лингвистических исследований*, 12 (3), 535-572.

Казаковская, Виктория Виладиевна; Ингрида Балчюниене; Лаура Камандулите-Мерфелдиене 2013, Цветовые прилагательные в речевом онтогенезе, *Проблемы онтолингвистики - 2013: Материалы международной научной конференции 26-29 июня 2013 г., Санкт-Петербург*, Санкт-Петербург: РГПУ им. А. И. Герцена, 132-137, <http://www.ontolingva.ru/ontolingv2013.pdf>.

Касевич, Вадим Борисович 1981, Слог в общем и дальневосточном языкознании, *Теория языка. Методы его исследования и преподавания. К 100-летию со дня рождения академика Л. В. Щербы*, под ред. Рубена Ивановича Аванесова, Ленинград: Наука, 141-146.

Кибрик, Андрей Александрович; Вера Исааковна Подлеская (ред.) 2009, *Рассказы о свидениях: Корпусное исследование устного дискурса*, Москва: Языки славянских культур.

Маркус, Даце; Алексас Гирденис 2011, Фонология дифтонгов [ie] и [uo] в речи латышских и литовских детей, *Онтолингвистика - наука XXI века. Материалы международной конференции, посвященной 20-летию кафедры детской речи РГПУ им. А. И. Герцена. 4-6 мая 2011 г., г. Санкт-Петербург*, 90-93, <http://www.ontolingva.ru/ontolingvistika2011.pdf>

Раева, Ольга Васильевна; Елена Игоревна Риехакайнен 2015, Синтагма и сегментация устной спонтанной речи, *Вестник Пермского университета. Российская и зарубежная филология*, 4 (32), 14-20.

Gauta: 2018-12-19

Priimta: 2019-01-14