

## **Determining Children’s Level of Acquisition through Grammatical Profiles: Evidence from a Bilingual Russian-English Child Acquiring Verbs**

*Maria Tsfasman*

In this paper, I present my research on verb acquisition of a bilingual Russian-English child. I suggest an alternative way of measuring the level of acquisition by determining the child’s grammatical profiles (i.e. the frequency distribution of word forms). I show how this method works on the material from one bilingual child, concluding that he has acquired the Russian verb system completely, making only few mistakes in verb formation. I also notice that grammatical profiles are different for spoken and written language.

In the first section I introduce the problem, describe terms and previous research on the topic (1.1), and present the data I used in my study (1.2). In section 2 I provide comparison between grammatical profiles of the child I studied (Jaša), the investigator that interacted with him during recording sessions, and corpus data analysed by Janda and Lyashevskaya (2011). Section 3 addresses mistakes Jaša makes while using verbs. In section 4 I draw conclusions, discuss the results and suggest areas for further research.

### *1. Introduction*

The main questions the current paper aims to answer are: what would grammatical profiles (the relative frequency distribution of different word forms) be like in a bilingual child, and how can they help us analyse a child’s level in language acquisition? I suggest a new method of analysing children’s proficiency in language through the distribution of forms they use and try to apply the method in a small-scale case study with data from one Russian-English bilingual child.

#### *1.1 Concepts and previous research related to the topic*

The emergence of Russian aspect in Russian-English bilinguals is well-studied by several authors, such as by Bar-Shalom and Zaretsky (2008), Gagarina (2007), and Pavlenko (2009). Thus, I did not aim to shed light on any particular area of aspect acquisition or bilingual specifics of this process in my study.



Linguistic profiling is a widely-used method, especially within Cognitive Linguistics and Construction Grammar. It includes radial category profiles, grammatical profiles, semantic profiles, constructional profiles, and collostructional profiles. All these methods are described in detail in Kuznetsova (2015). The main idea behind linguistic profiling is to show that there is a strong connection between form and meaning, and that by studying form distribution it is possible to discover something new about the meaning.

In the current work, I only use grammatical profiling, which, first introduced by Janda and Lyashevskaya (2011), is a method predicting a word's behaviour through the relative frequency distribution of the inflected forms of a word.

The stage in a child's language development is usually determined by mean length of utterance (known as MLU); the proportion of mistakes they make in their speech, a system is considered fully acquired when the percentage of mistakes equals at 90 and more percent; or various tests, such as Multilingual assessment instrument for narratives (see Gagarina et al. 2012), and British picture vocabulary test (see Dunn L. & Dunn D. 2009). However, all those procedures include manual work and cannot be fully automated, while the method I propose can.

Grammatical profiling has not been applied to child acquisition of language in previous studies. However, a related method was used by O. Laleko in "What Errors Can't Tell Us About Heritage Grammars: On Covert Restructuring of Aspect in Russian" (2010). This research addresses the differences in distribution of the aspects in adult heritage speakers and monolinguals. She concludes that monolingual distribution of aspectual pairs is different from that of heritage speakers. This suggests that it is reasonable to observe a child's grammatical profiles, and that it would be also reasonable to expect substantial differences from adult speakers.

### *1.2 Data and method*

My research is based on data collected by Tania Ionin, presented in CHILDES (<http://childes.psy.cmu.edu/access/Biling/Ionin.html>). I used 2 transcripts of taped interaction with a bilingual Russian-English child, Jaša. Surprisingly, this material was used only for studies of the acquisition of English (such as Ionin 2001, 2002, 2008), but not of the acquisition of Russian, which makes me pioneer in analysing this part of the data.

Jaša was born in the USA in 1992 to parents who had emigrated from Russia. Russian for him was the primary language spoken at home by his parents and grandmother. At the time of the first recording Jaša was 4;7 years old, the second one is from one month after (Jaša was 4;8).

The transcripts do not contain any morphological tags; thus, all verb utterances were extracted manually by reading through every transcript. (I did not use any programs that automatically determine part of speech and concrete form, as all the scripts are transliterated in an inconsistent manner; in future research my method can be fully automatized.) Analysed data consisted of 825 verb utterances in total (see table 1 for exact numbers of verb forms and verbs).

	Jaša	Investigator	Total
Verb utterances	495	330	825
Verbs	142	111	184

Table 1: Number of verb forms and verbs in my data

For each example, I wrote down what verb form it is, what its corresponding infinitive is, and tagged their morphological characteristics: aspect, tense, number and person if a verb was non-past, number and gender if it was in the past tense, number if it was imperative.

Further, I compare data from Jaša with the material from the Russian National Corpus presented and analysed in Janda and Lyashevskaya (2011) in order to identify differences and similarities between the child's verb system and an average adult Russian language speaker. Janda and Lyashevskaya's data consisted of 5,951,250 verb forms and was thoroughly analysed applying various statistical methods. I also compare Jaša's speech with the investigator to avoid methodological mistakes related to the difference between written (corpus data) and spoken language (which are represented by the corpus data and mine correspondingly), style and so on. I make the assumption that if Jaša's verb system is similar to the investigator's, the system can be considered target-like, even if it has considerable differences with the data from corpus. This method can be criticised because the investigator's speech is child-directed and its morphological richness can be affected by that (see Xantos et al. 2011). Thus, in further research it would be reasonable to compare children data with oral corpus material.

To see to what extent Jaša has acquired verbs at the time of the recordings I use statistical tests and analyse the data in R (R Core Team 2017, Wickham 2009, Hothorn & Zeileis 2015).

## 2. Grammatical profiles

The first thing that comes to mind when conducting an analysis on language acquisition is to compare data from children with data from adult native speakers. In this way, it is possible to understand how children's speech is different from the speech of linguistically fully proficient speakers.

In this section, I argue that Jaša's verbs are adult-like in their distribution and in relation to their grammatical profiles. I compare data from Jaša with the material from the Russian National Corpus presented and analysed in Janda and Lyashevskaya (2011). Moreover, in this section I provide a comparison of Jaša's verbs with the distribution of verbs used by the investigator while interacting with Jaša from the same transcripts. This analysis aims at verifying whether the differences between the data from Jaša and the corpus come from differences in genres and type of texts – or reflect the actual level of Jaša's language.

### 2.1 Jaša's choice of verb forms

First, I will describe the distribution of all verb forms and how Jaša's grammatical profiles are different from the investigator and a Russian language user in general (which I assume is what corpus data shows).

To illustrate the distribution, I present bar charts from Janda & Lyashevskaya's (2011) (fig. 1) and similarly structured charts for data from Jaša (fig. 2) and the investigator (fig. 3).

In this subsection, I will compare only distributions of concrete verb forms in past and non-past tenses, while in the next subsections I turn to the distribution of forms depending on aspect. See appendix 1 for the raw numbers used in constructing plots for Jaša and the investigator.

The graph in figure 1 is somewhat different from those in 2 and 3 – on the y-axis it shows percentages of forms from the total number of perfective and imperfective verb forms separately. In figures 2 and 3 the percentage is counted from the whole number of verb forms for a speaker not regarding verbs aspect.

In all 3 plots the x-axis shows what tense and aspect the percentage is for. Colouring is for different concrete forms.

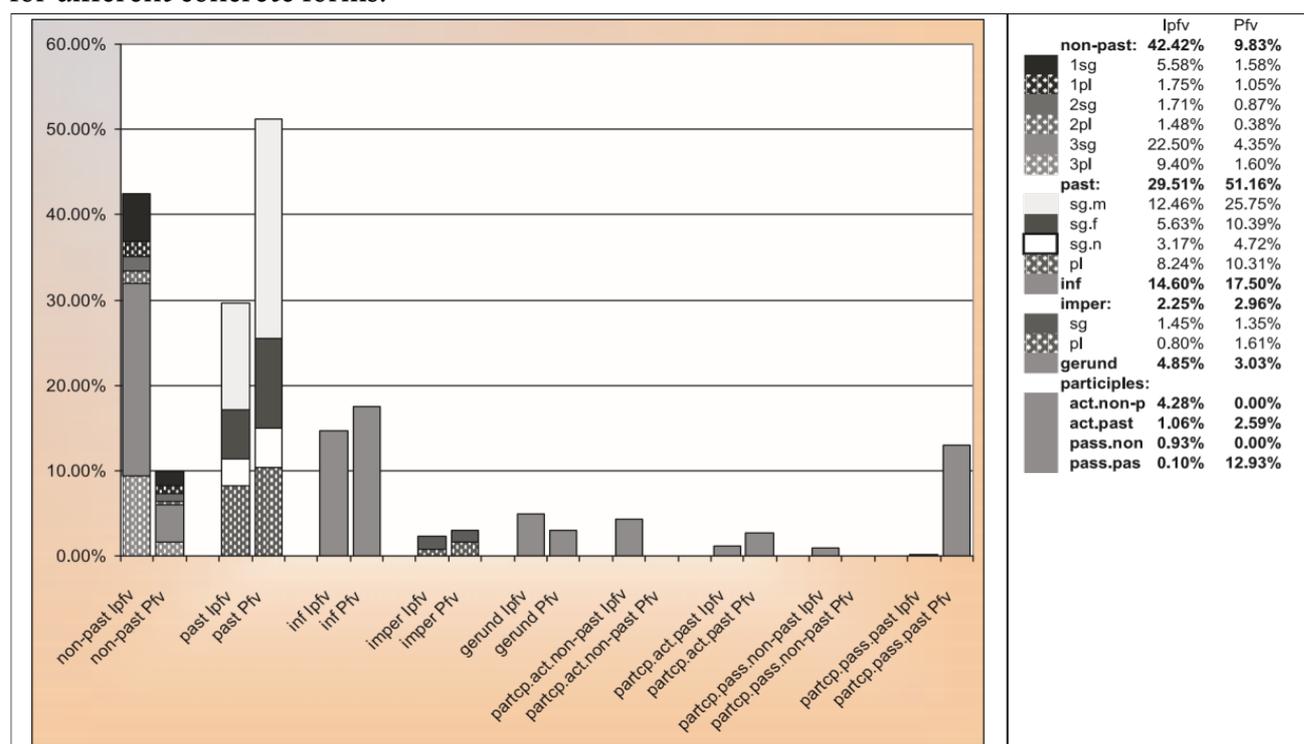


Figure 1: The distribution of different verb forms in corpus data (from Janda & Lyashevskaya 2011)

The figures show that within verbs in the non-past tense Jaša uses much more 1sg (about 28%) than the investigator (only 5%), while the investigator has a higher percentage of 2sg than Jaša (15% and 4% respectively). However, corpus data shows a higher rate of 3sg (in corpus data about 30% of all verb utterances are 3sg, while for Jaša and the investigator the proportions of such forms estimate at about 22%). This shows that grammatical profiles strongly depend on context and the situation language is used in. The investigator's aim is to encourage Jaša to speak, so she asks many questions and talks mostly about Jaša's life; as a result, she uses a large number of verbs in the 2<sup>nd</sup> person. Jaša, however, tells the investigator about himself and his life, which makes him use a lot of verbs in the 1<sup>st</sup> person. On the other hand, corpus data represents written language, which usually describes characters in the 3<sup>rd</sup> person.

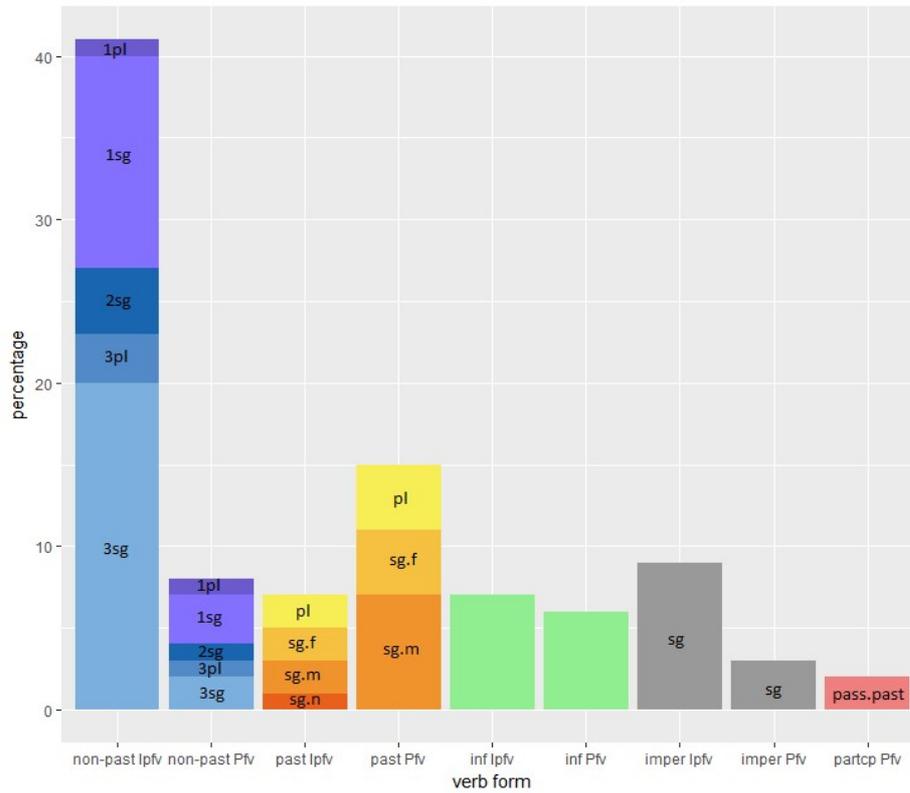


Figure 2: The distribution of verb forms in Jaša's speech

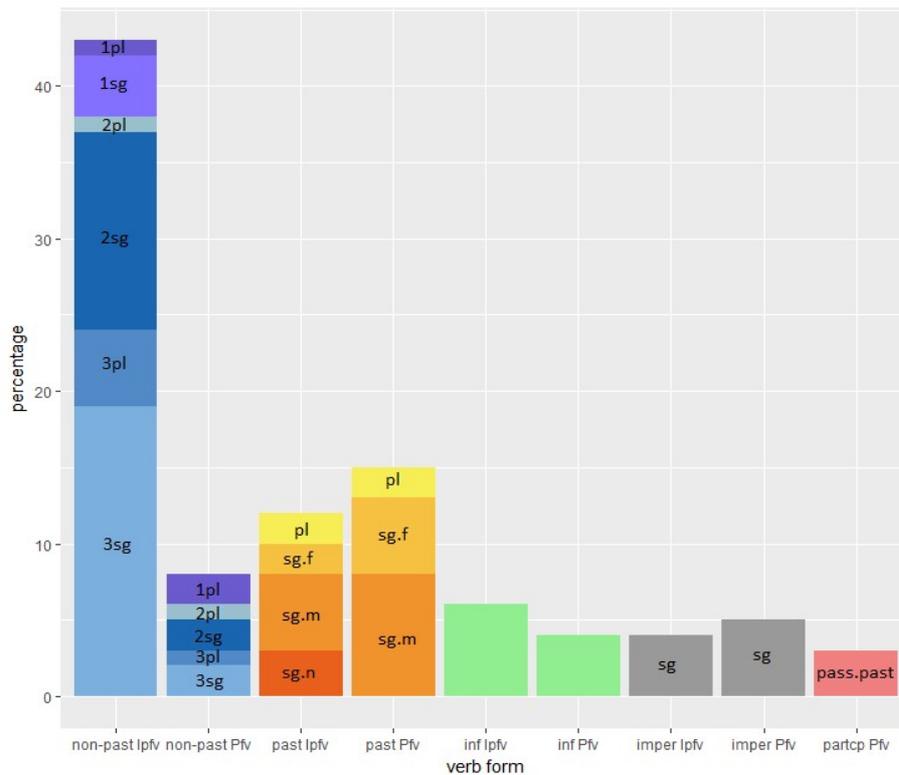


Figure 3: The distribution of verb forms in the investigator's speech

However, the distribution of past tense verb forms (different by gender and number) is more or less similar for Jaša, the investigator and data from the corpus. Slight differences are likely to be due to the limited data I have.

To sum up, Jaša's grammatical profiles strongly resemble those of adult native speakers. Those differences that appear between Jaša's and corpus non-past tense verb form distribution are explainable by the diversity of texts used in those datasets (written vs. spoken language). Differences between Jaša's and the investigator's distributions are likely to come from different purposes of their speech.

## 2.2 Jaša's choice of aspect

Janda and Lyashevskaya (2011) demonstrated in their paper that grammatical profiles for aspectual pairs are significantly different (see table 2). They argue that imperfective verbs are more frequently used in non-past tenses, while perfectives much more often tend to be used in past tenses. To answer the question of whether Jaša had acquired aspect and their grammatical profiles I made similarly structured tables for Jaša's (table 3) and the investigator's (table 4) verb distribution.

All tables show the percentage of different forms from the total number of imperfective and perfective verbs separately.

	Imperfective				Perfective			
	Ipfv_ NonPast	Ipfv_ Past	Ipfv_ Inf	Ipfv_ Imper	Pfv_ NonPast	Pfv_ Past	Pfv_ Inf	Pfv_ Imper
both p- & s-partners	1,330,016 47.4%	915,374 32.6%	482,860 17.2%	75,717 2.7%	375,170 11.9%	1,972,287 62.7%	688,317 21.9%	111,509 3.5%

Table 2: Grammatical profiles of imperfective vs. perfective verbs from corpus data (from Janda & Lyashevskaya 2011)

	Imperfective				Perfective			
	Ipfv_ NonPast	Ipfv_ Past	Ipfv_ Inf	Ipfv_ Imper	Pfv_ NonPast	Pfv_ Past	Pfv_ Inf	Pfv_ Imper
number	208	33	33	47	36	78	30	17
percent	65%	10%	10%	15%	22%	48%	19%	11%

Table 3: Grammatical profiles of imperfective vs. perfective verbs for Jaša

	Imperfective				Perfective			
	Ipfv_ NonPast	Ipfv_ Past	Ipfv_ Inf	Ipfv_ Imper	Pfv_ NonPast	Pfv_ Past	Pfv_ Inf	Pfv_ Imper
number	141	39	19	12	29	43	14	18
percent	67%	18%	9%	6%	28%	41%	13%	17%

Table 4: Grammatical profiles of imperfective vs perfective verbs for the investigator

First, I will compare Jaša's grammatical profiles to those of the investigator. Tables 3 and 4 show that the distribution of forms is mostly similar: imperfective verbs are in most cases used in the non-past tense (65% for Jaša and 67% for the investigator), and the percentages of the infinitives in imperfective verbs are also the same. Furthermore, most of the

perfective verbs for both speakers are used in the past tense, and the second most used form in perfective verbs is the non-past tense. Finally, the rates of perfective infinitives are also nearly the same. However, Jaša uses much more imperfective imperatives (15%) than the investigator (6%), and he uses them more than perfective imperatives (11%). This difference is caused by one outlier – the verb form *smotri* (eng. look). Jaša draws the investigator’s attention to various objects by saying “Look!” (*Smotri!*) many times during both analysed transcripts. For this reason, the observed difference cannot be considered representative. Based on my findings, I conclude that Jaša’s aspectual grammatical profiles are adult-like, and that he has acquired the aspectual system on a highly proficient level.

However, the distributions in my data and corpus data are different: both Jaša and the investigator use a larger proportion of non-past tense in verbs of both aspects (around 65% in perfective and 25% in imperfective) than I find in the corpus data (47% in perfective and 12% in imperfective). This is most likely due to the genre – it is more common to use past tense in written language and less common to use it while speaking. What is more, the percentage of imperfective is also higher in my data than in the data from the corpus. The reason is probably the same – in spoken discourse and dialogue imperfective is more common than in the written language. This comparison suggests that grammatical profiles are different in spoken and written language. Thus, it would be interesting to compare Janda & Lyashevskaya ‘s (2011) data to data from an oral corpus, although this is a task that is beyond the scope of the present study.

### 2.3 What factors really matter: CART analysis

To find out what factors are relevant and how they interact I conducted a Classification and Regression Tree (CART) analysis presented in figure 4, and a variable importance plot based on Random Forest analysis showed in figure 5 (R Core Team 2017, Hothorn & Zeileis 2015).

Building a classification tree is a way to visualize decision rules for predicting a categorical outcome. The model goes through the whole data and shows which factors affect the distribution of a dependent variable. In my analysis, aspect was chosen as the dependent variable and verb form (their concrete characteristics, such as tense) and speaker – independent variables. I built a CART to show whether Jaša’s aspectual grammatical profiles is different from the investigator’s.

As can be seen from figure 4, what affects the choice of aspect the most is verb form. This fact indicates the existence of different grammatical profiles for different aspects. Speakers affect the choice only when the verb is in the imperative, which, as I mentioned above is caused by a single outlier (*smotri* ‘look’) in Jaša’s speech. What is more, the p-value for the speaker factor is relatively high, which is yet another indication that this factor is not highly relevant. To summarize, the CART analysis in figure 4 serves as another indication that Jaša’s grammatical profiles do not differ from those of an adult native speaker.

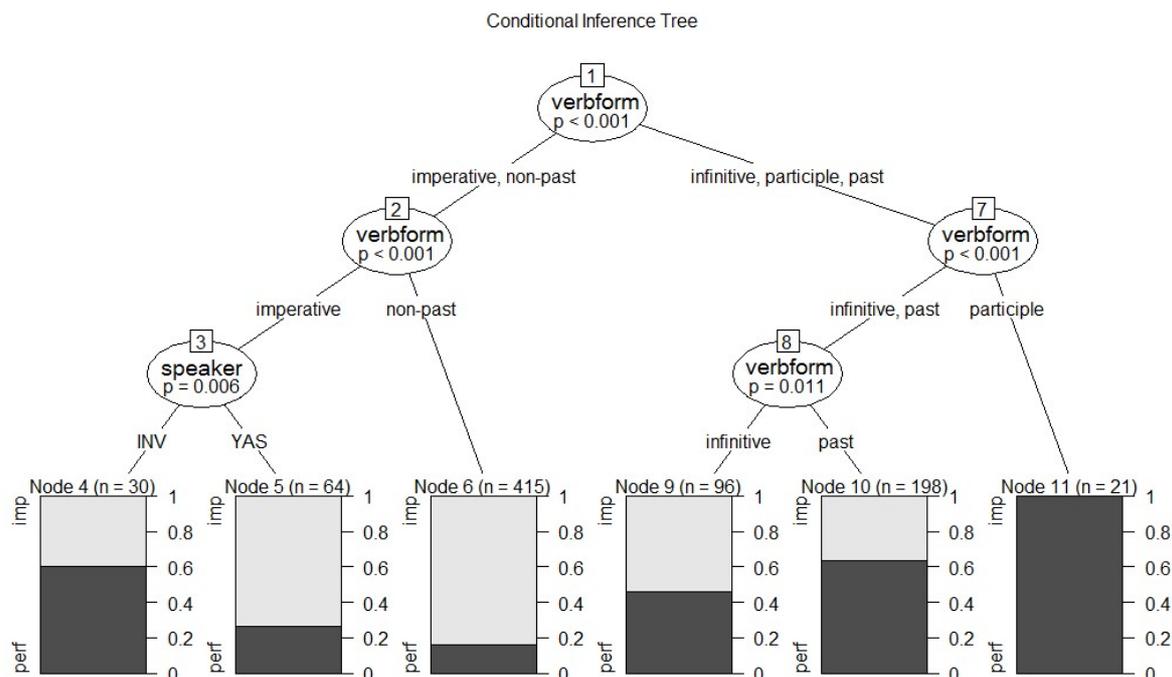


Figure 4: Conditional inference tree for the whole data

### 3. Analysis of mistakes

In section 2 I hypothesized that Jaša's verb system is complete and adult-like, according to his grammatical profiles. Since traditionally such conclusions are made based on analysis of mistakes, here I will provide such an analysis for Jaša to show that my results correspond to those of a well-established method. The fact that in 92,5 % of the examples he used correct forms shows that he had reached target-like verb morphology, which is expected from a child of his age. However, as we will see, he still makes some mistakes that a typical adult Russian speaker would not make.

To begin with, I will describe what kinds of mistakes Jaša makes (for a detailed table with all Jaša's mistakes, see appendix 2). I divided incorrect verb forms in 5 groups shown in table 5 with the percentage of mistakes of certain types from all Jaša's verb utterances.

Type of mistake	Raw numbers	Per cent
Phonological	19	3,8%
Wrong stem	6	1,2%
Mistake in prefix	3	0,6%
Wrong aspect	5	1,0%
Wrong verb form	3	0,6%
Total	37	7,5%

Table 5: Mistakes frequency by type (per cent from the total number of examples from Jaša)

Most of Jaša's mistakes are phonological and are caused by his phonological development – for example, he mispronounces verbs with the [r] sound such as *nasovat'* instead of *narisovat'* 'draw', *govili* instead of *govorit'* 'talk', *otkivat'* instead of *otkryvat'* 'open'. He also sometimes swallows the last syllable, for instance, *bi* instead of *byli* 'was'. More interesting for me are mistakes of the 4 other types.

In 6 cases Jaša used an incorrect stem form. He either uses non-past stem in a past verb form such as *voz'měl* instead of *vzjal* 'took' or *raskrašit'* instead of *raskrasit'* 'paint'. Or he uses the wrong stem alternant within the non-past tense subparadigm – *posadju* instead of *posażu* 'I will plant'.

Although Jaša's aspectual system is adult-like, in 5 cases he uses aspects not suitable in the relevant context. He both confuses imperfective with perfective (in two cases as in ex. 1) and the other way around (in 3 cases as in ex. 2).

- (1) Net, net, net, ne zaberil! (instead of imperfective *ne zabiraj*)  
'No, no, no, don't take it away!
- (2) Èta bašnja mozet padat'! (instead of perfective *upast'*)  
'This tower can fall!'

The rarest mistakes are in prefixes – he either uses an extra prefix when not needed (3) or he does not use prefix when the verb cannot be used without it (4):

- (3) Kogda Gulja narodilas' (instead of unprefixes *rodilas'*)  
'When Gulja was born'
- (4) A potom mašinka malen'kaja ego dvinet (instead of prefixed *sdvinet* or *podvinet*).  
'And then a little car will move it'

When it comes to the category "wrong verb form", in 2 examples Jaša used a singular instead of a plural verb form, and one time he used infinitive instead of a finite form.

To sum up, Jaša makes some mistakes that one would not expect from a proficient adult native speaker. Nevertheless, most of these mistakes are not morphological and in general the percentage of incorrect forms is very low, so Jaša's use of verbs is very close to that of adult language users.

Thus, a traditional method leads me to the same conclusions as my analysis of grammatical profiles. However, while grammatical profiles can be tested automatically, analysis of mistakes requires manual data tagging. More importantly, mistakes can show what a child misuses, but do not enable us to spot which forms a child avoids, which is also important when it comes to language acquisition.

#### 4. Conclusions and discussion

Grammatical profiles in children seems to be a good and straightforward way to see on what level a child has acquired the grammar. Jaša, a bilingual Russian – English speaker with Russian as a minority language, shows proficiency in Russian verb usage at 4 years 7-8 months. He makes phonological and morphological mistakes in only 7,5% of verb utterances, and his grammatical profiles are adult-like.

The proposed alternative method for determining child's linguistic level could be easy to automatize, and with present day child speech recording technology it would be easy to create large sets of data and, maybe, discover more discrete developmental stages. However, it is important to note that my method requires more evidence from a larger set of data. First, to make the picture complete it would be relevant to compare my data with

the verb usage in the oral subcorpus of the Russian National Corpus. Second, in further research, it is essential to compare grammatical profiles of children at different stages of their linguistic development.

Moreover, my research gives an indication that grammatical profiles are different for perfective and imperfective verbs. In this paper, I show that grammatical profiles are different depending on what type of texts they are based on. When studying grammatical profiles, it appears crucial to take into account the type of discourse and genre, and even the type of situation may matter.

### References

- Bar-Shalom, E., & Zaretsky, E. (2008). Selective attrition in Russian-English bilingual children: Preservation of grammatical aspect. *International Journal of Bilingualism* 12(4), 281- 302. <http://doi.org/10.1177/1367006908098572>
- Dunn, L. M., & Dunn, D. M. (2009). The British picture vocabulary scale. Place of publication not identified: GL Assessment Limited.
- Gagarina, N. V., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U. & Walters, J. (2012). MAIN: Multilingual assessment instrument for narratives. Berlin: Zentrum für Allgemeine Sprachwissenschaft.
- Gagarina, N., Armon-Lotem, Sh. and Gupol, O. (2007). Developmental variation in the acquisition of L1 Russian verb inflection by monolinguals and bilinguals. *BUCLD 31 Online Proceedings Supplement*: <http://www.bu.edu/buclid/proceedings/supplement/vol31/>
- Ionin, T. (2008). “Progressive aspect in child L2-English”. In B. Haznedar and E. Gavruseva (Eds.), *Current Trends in Child Second Language Acquisition: A Generative Perspective*. John Benjamins, pp. 17-53. <http://doi.org/10.1075/lald.46.03ion>
- Ionin, T. and Wexler, K. (2001). “L1-Russian children learning English: Tense and overgeneration of ‘be’.” In X. Bonch-Bruevich et al. (Eds.), *Selected proceedings of the 2000 Second Language Research Forum* (pp. 76-94). Somerville, MA: Cascadilla Press.
- Ionin, T. and Wexler, K. (2002). “Why is ‘is’ easier than ‘-s’?: Acquisition of tense/agreement morphology by child L2-English learners.” *Second Language Research* 18, 95-136. <http://doi.org/10.1191/0267658302sr195oa>
- Janda, L. and Lyashevskaya, O. (2011). Grammatical profiles and the interaction of the lexicon with aspect, tense and mood in Russian. *Cognitive Linguistics* 22:4, 719-763. <http://doi.org/10.1515/cogl.2011.027>
- Kuznetsova, J. (2015). *Linguistic Profiles*. Berlin/Boston: De Gruyter. <http://doi.org/10.1515/9783110361858>
- Laleko, O. (2010). What Errors Can’t Tell Us About Heritage Grammars: On Covert Restructuring of Aspect in Russian. *LSA Annual Meeting Extended Abstracts* 1, 52: 1-5.
- Pavlenko, A. (2009). Verbs of motion in L1 Russian of Russian-English bilinguals. *Bilingualism: Language and Cognition* 13(01), p.49. <http://doi.org/10.1017/S1366728909990198>
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.r-project.org/project.org/>
- Torsten Hothorn, Achim Zeileis (2015). partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research* 16, 3905-3909. URL

<http://jmlr.org/papers/v16/hothorn15a.html>

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

<http://doi.org/10.1007/978-0-387-98141-3>

Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., Gagarina, N., Hrzica, G., Ketrez, F., Kilani-Schoch, M., Korecky-Kröll, K., Kovačević, M., Laalo, K., Palmović, M., Pfeiler, B., Voeikova, M. and Dressler, W. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Language* 31(4), pp.461-479. <http://doi.org/10.1177/0142723711409976>

*author: Maria Tsfasman*

*affiliation: National Research University – Higher School of Economics,  
Moscow*

*email: mmtsfasman@gmail.com*