Russian yers: Child language data as a window on regularity and irregularity

One of the most widely studied sound patterns in Russian involves vowel-zero alternations, referred to as jers/yers (e.g., Bethin 1998, Gouskova & Becker 2013, Itkin 2007, Lightner 1972, Rubach 1984, Sheer 2010, Yearley 1995). In the canonical case, an unstressed /o/ or /e/ is present in some inflected forms, but absent in others. Alternations of this kind would appear to be a perfect test case for how native speakers learn to reconcile lexical idiosyncrasies with productive generalizations. Yet, there is, to date, little work that directly measures how Russian-speaking children produce these alternations across multiple lexemes and morphosyntactic environments in a controlled elicitation setting. In this study, child language data is elicited in an attempt to determine: (i) whether deviations from adult forms are concentrated in particular lexemes or in particular grammatical environments, and (ii) whether these deviations are best stated in terms of phonological, morphological, and/or morphosyntactic features. The broader goal is to use child production data to build realistic and constrained theories of morphophonological learning.

The study uses a production elicitation experiment with a monolingual population of Russian preschoolers attending a public kindergarten in St. Petersburg, Russia (N=34, ages 2;11–5;01). Each picture stimulus was paired with 2–4 questions targeting distinct inflected forms of the same lexeme, including both vowel-full and vowel-less contexts (see Table 1). Target items are mono-, bi-, and tri-syllabic nouns, including derived forms with high-frequency affixes like *-ok* DIMINUTIVE (e.g., *kusók* 'piece') and potentially idiosyncratic ones (e.g., *ogón*' 'fire'). Nouns with vowel-zero alternations (e.g., *botínok* 'boot') and without (e.g., *sapóg* 'high boot') were included. In comparing children's productions to adult forms, coding of mismatches included the following variables: (a) diminutive (a diminutive form is used, e.g., *ugolók* 'corner.DIM' instead of *úgol* 'corner'), (b) child vowel (a vowel is realized where none is expected in the adult grammar), (c) no vowel (a vowel expected in the adult form is omitted by the child), (d) case/number mismatch, (e) stress deviation, (f) overgeneralization pattern.

Table 1: Sample stimuli to elicit different forms of *ugol* 'corner'

Stimulus	Question	Target lexeme	Case
	*Мальчик себя плохо вел, куда его поставили родители?* 'The boy misbehaved, where did his parents put him?'	(v 'in') úgol 'corner'	ACC
	Где он теперь стоит? 'Where is he standing now?'	(v 'in') uglú 'corner'	LOC

Table 2 presents preliminary results with illustrative examples. Lexemes were also mapped to fine-grained declension categories (Zaliznyak 1967, 1977), revealing that declension type 3*a/b resulted in fewer alternation mismatches, while 1/2*b lexemes were most challenging.

Table 2: Preliminary results

Diminutives show lower rates of mismatch than corresponding non-diminutives.	*ogon'ók* 'fire.DIM' (1.00) vs. *ogón'* 'fire' (0.83); *pen'óček* 'stump.DIM.DIM' (1.00) vs. *pen'ók* 'stump.DIM' (0.94) vs. *pen'* 'stump' (0.75)
Among non-diminutives, some lexemes show higher rates of mismatch than others. Further, these lexical effects explain more variance than children's individual differences.	*lev* 'lion' (0.66), *kost'ór* 'campfire' (0.68), *úgol* 'corner' (0.95), *cypl'ónok* 'chickling' (1.00) SD: Participant=1.09, Lexeme=1.91 ICC: Participant=0.15, Lexeme=0.45

Mismatches involving overgeneralization suggest productive or analogical use of vowel-zero alternation in /-ok/ contexts.	*sapóg* ‘boot.NOM.SG’ – **sapká* ‘boot.GEN.SG’ (N=2), *ut’ónok* ‘duckling.NOM.SG’ – **ut’ónki* ‘duckling.NOM.PL’ (N=3)

Table 3 summarizes key findings by linguistic category and subcategory. Interestingly, stress mismatches were not correlated with any other mismatch types.

Table 3: Summary of key findings by category

Category	Subcategory	Details
Phonetic/ phonological	Stress	Last stem vowel stress preferred: **nóžem* for *nožóm* (NOM.SG nož ‘knife’) (N=9), **sapóga* for *sapogá* (NOM.SG *sapóg* ‘high boot’) (N=4); e.g. child_1: *červ’áka* ‘worm.GEN.SG’, *sapóga* ‘high boot.GEN.SG’, *nóžem* ‘knife.INST.SG’, *lóbe* ‘PREP.SG’
Phonetic/ phonological	Clusters	Children often use vowel-less preposition forms in phonological words with a *-dl’d-* medial and *sl’v-* initial consonant clusters when both vowel-less and vowel-full forms are possible: *pod l’dóm* (N=6) vs. *podo l’dóm* (N=2) ‘under ice’, *s l’vom* (N=13) vs. *so l’vom* (N=2) ‘with lion’
Morphological/ lexical	Diminutives	Overall, adult vowel-zero alternations are more frequent in diminutives than non-diminutives. However, some diminutives are less transparent than others (for lexemes *lev* ‘lion’ and *kost’ór* ‘campfire’, diminutives were never used).
Morphosyntactic	Case/number	Syncretic case forms consistently result in fewer alternation mismatches ($\beta = -2.22$, $p < .001$).
Morphosyntactic	Case/number	Frequent case mismatches include PREP.SG for LOC.SG, NOM.SG for GEN.SG and ACC.SG, GEN.PL for GEN.SG
Morphosemantic	Animacy	In accusative case, children favor animate forms when animate or inanimate are possible: *igrat’ v kúkolok* ‘to play dolls.ACC.ANIM’

For computational modeling, lexeme frequencies were estimated from two sources: combined texts of children’s books used in the classroom and child-directed speech from CHILDES corpora (Lelik et al. 2025). These estimates serve as a proxy for input type and token frequencies, computed separately for vowel-full and vowel-less realizations of each target lexeme. Following Jarosz’s scaled-weights MaxEnt approach (Goldwater & Johnson 2003, Jarosz 2025), a lexically conditioned learner was trained on these per-lexeme proxy distributions. Preliminary results of this model show that the model’s predicted P(vowel-less) is close to children’s performance for some lexemes, but diverges for others, with both under- and over-prediction.

Future work will extend this comparison to additional learning models, including the Lexicalized Expectation Driven Parameter Learner (Jarosz et al. 2024), which incorporates a distinct lexical component. It will also incorporate other frequency proxies to maximize/simulate relevance to the actual input, such as child-directed data from kindergarten classes, children’s songs, nursery rhymes and games.

References

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press, Cambridge.

Bethin, C. Y. (1998). *Slavic Prosody*. Cambridge: Cambridge University Press.

Goldwater, S. & Johnson, M. (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University.

Gouskova, M., & Becker, M. (2013). Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory*, *31*(3), 735–765.

Itkin, V. A. (2007). *Russian Morphonology* [*Russkaya morfonologiya*]. Moscow.

Jarosz, G., Hughes, C., Lamont, A., Prickett, B., Baird, M., Kim, S. & Nelson, M. (2024). Type and Token Frequency Jointly Drive Learning of Morphology. OSF.

Jarosz, G. To appear. Incremental Learning of Lexically-Specific Morphophonology: an Integrative Approach. To appear in *Linguistics Vanguard*.

Lelik, V., Diachkova, M., Dorofeeva, S. V., Lopukhina, A., Dragoy, O., Sekerina, I. A. (2025). RusLan-M Corpus. Retrieved from https://childes.talkbank.org

Lightner, T. M. (1972). *Problems in the Theory of Phonology, Vol. I: Russian Phonology and Turkish Phonology*. Edmonton: Linguistic Research, Inc.

Rubach, J. (1984). *Cyclic and Lexical Phonology: The Structure of Polish*. Foris, Dordrecht.

Rubach, J. (2013). Exceptional segments in Polish. *Natural Language and Linguistic Theory* 31: 1139–1162.

Sheer, T. (2010). Why Russian vowel-zero alternations are not different, and why Lower is correct. *Journal of Language and Verbal Behaviour* 9: 77–112.

Yearley, J. (1995). Jer Vowels in Russian. In J. N. Beckman, L. W. Dickey and S. Urbanczyk, eds., *Papers in Optimality Theory*. Amherst: GLSA: 533–571.

Zaliznyak, A. A. (1967). *Russian Nominal Inflection* [*Russkoe imennoye slovoizmenenie*]. Nauka, Moscow.

Zaliznyak, A. A. (1977). *Grammatical Dictionary of the Russian Language* [*Grammaticheskiy slovar' russkogo yazyka*]. Russky Yazyk, Moscow.