

Writing-based wordforms vs. spoken wordforms

Vadim Kasevich¹, Iuliia Menshikova²

¹Faculty of Asian and African Studies, SpbU, Russia

²Faculty of Philology, SpbU, Russia

<https://doi.org/10.36505/ExLing-2016/07/0015/000274>

Abstract

This study addresses a vital problem of reshaping Russian Grammar in conformity with its real acoustic realization rather than with the traditionally written expression plane. In this way, one can switch from an absolutely abstract coding of wordforms to acoustic entities, first phonological and then phonetic, which underlie the real processes of speech production and speech perception. The multiple approach to grammar writing makes it necessary to develop a special database for the phonologically represented wordforms of Russian. Typically, the respective paradigms are reduced. More generally, the links camouflaged by the traditional orthography are made visible. E.g., the Adjective Gender paradigm, normally made up of three genders, is reduced to a two-item paradigmatic structure, because Neuter Gender and Feminine Gender just merge.

Key words: linguistics, Russian language, grammar, phonetics, morphemics.

Introduction

Natural language grammars as may differ in many ways depending on the theories that underlie them. However different, the vast majority of grammars share at least three important things in common, viz. (i) practically all of them are designed to account for the formal structure of the language rather than for its functioning, (ii) even where the grammars somehow model the dynamic nature of the language, the sets of rules are typically intended for the speakers rather than for the hearers, (iii) most grammars present their paradigms etc. in terms of standard orthography rather than in terms of phonological representations.

Unlike the prevailing tradition referred to above, we choose an approach where the grammar (of Russian) is modelled as a set of rules designed for the hearer. Since the hearers operate with sound patterns of linguistic entities the expression plane of the entities is expected to be presented in terms of the phonemes. E.g. the wordform КУПАТЬСЯ (kupaťs'a) 'bathe' is normally written with the so-called particle –СЯ (–s'a). However, if we switch to its sound shape, we find that the hearer must be prepared to recognize, in addition to /kupaľ-s'a/ 'bathed', also /kupaľ'i-s'/ '[they] bathed', and /kupać-ca/ '[to] bathe'. In many cases, the phonology-based representation reshapes the paradigm as compared with its writing-based version, cf. HOBOE Neuter 'new' and HOBAЯ 'new' Feminine which just merge in /novaja/.

ExLing 2016: Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics, 27 June – 2 July 2016, Saint Petersburg, Russia

Methods

As our goal is to “redress” Russian morphology in such a way that its expression plane would be consonant with the phonology, our first step is to provide all the (nominal) wordforms of the Russian lexicon with a phonological transcription. E.g., ОДЕЯЛ-О ‘blanket’ → /ad’ijál-a/. To make sure that our phonological transcription faithfully reproduces the expression plane of the wordforms chosen, our Ss. were asked to filter out the output of the transcribing routine.

The second step is developing a database for nouns where all the relevant information about individual nouns would be stored (see Kasevich *et al.*, this volume).

Results and discussion

The results of reshaping Russian nominal wordforms along phonological lines make it possible to see the morphological structure of Russian ‘as it is’, with a “distorting” influence of the traditional writing totally eliminated.

For instance, it is a well known fact that in many, if not all, languages where morphological component is sufficiently developed, the paradigms include at least two homomorphic inflections, cf. DOM ‘house’, Nominative and DOM ‘house’ Accusative. (One could add, rather parenthetically, that if such pairs would be the only means to express given meanings, there would be every reason to classify Russian with Ergative languages.) When we base our analysis on spoken (phonologically represented) wordforms, two more homomorphic forms should be added to the DOM-paradigm, viz. /dom-i/ ‘house’ Genitive and /dom-i/ ‘house’ Locative. Using our database, one can easily trace all the types of paradigm reduction due to the spoken-form orientated approach. What is more important, in this way we can try to bring to light the regularities that underlie the functioning of the grammar. For instance, we can see that Neuter is a ‘weak’ point of the paradigm it enters, as it tends to merge with Feminine (cf. above).

We are not going to claim that the traditional writing based grammars are just “cultural artifacts” with no prototype in the real world. However, we do claim that spoken language should be given priority, if one sets an ambitious goal of looking into inner mechanism of language. That would be consonant with the insights from linguists like Jan Baudouin de Courtenay, Lev Scherba and Charles Hockett who insisted on an absolute necessity to discriminate between differently aimed grammars.

A typological note would be appropriate. For quite a few languages, the problems discussed in this paper are simply irrelevant, because the

languages are pre-literate. As a matter of fact, compiling special Russian grammars intended for the hearers treats the Russian language as if it were pre-literate. Another situation is met where there is a wide gap between writing and sound systems. If we compare, say, Russian and English, we will see that the Russian writing system is relatively simple and systematic, while the English system is notorious for its very unsystematic, sometime extravagant, relationship between writing and sound. This means that the analyst will be confronted with very different tasks depending on the language.

It is also interesting to study the sound-writing relation from the point of view of how writing reflects diachronic shifts. For Russian, it could be hypothesized that, at least in some cases, the reduction phenomena described above synchronically recapitulate diachronically important development (like Weak Vowel Drop, etc.).

Finally, a few more words about our problem from the applied linguistics perspective could be added. Stripping the wordform of its writing 'dress' is not the end of the story, although it is surely a prerequisite to writing computer programs for automatic speech perception and speech production. A phonologically transcribed speech, especially when it is a piece of the fluent text, is still very far from the real acoustic speech signal with all its redundancy on the one hand and imperfections and missing portions on the other. It is quite typical to be exposed to a speech signal so impoverished that only a good deal of guesswork makes an adequate perception possible.

There is one more very important problem that cannot be neglected, given the goal of our study. We mean the prosodic (here accentual) characteristics which are indispensable for any wordform of Russian. It has been demonstrated in lots of experiments that the lexical stress (accent) is an independent parameter in speech perception. According to our findings, quite typical is the situation where accent recognition scores are much higher than those for the phonemes or syllables. It is much likely that the overall language system contains a separate, relatively independent prosodic subsystem. This subsystem comes into play first in speech perception and in language acquisition, too, the stress strategies are well developed even prior to all the other subsystems.

Here again, typological aspects are also essential. To begin with, there exist languages, like Mongolian, where they have no lexical accent (stress) at all (vowel harmony being a partial functional substitute). No statistics are available, but it seems safe to argue that the number of unaccentual (lacking lexical stress) languages are much less. However, if we turn to

standard written texts, where no accents are shown, we will see that the two language types discussed above (with and without stress) become very much closer. Within one language as well as cross-linguistically, various subsystems and compensatory strategies are used to achieve an approximately the same level of efficiency both in perception and production, writing being one of the factors in play.

Writing to some extent makes obscure the real number of the homonyms to be found in the language. According to our data, in Russian one finds more than four thousand words which are written the same but differs due to different positions of the stressed syllable, e.g. L'UBIM ~ L'UBIM' '[we] love ~ '[he is] loved'. These are, so to speak, writing-made homonyms although 'in reality' they are a clear case of minimal pairs.

In some cases, the writing/spoken dichotomy may determine the very deep typological features making the language typologically the way it is.

According to a witty observation of Professor EugenyJakhontov, Semitic languages are typologically close to the isolating class when the languages are written, but acquire most features of inflexional languages when the languages are spoken, The thing is that in Semitic languages the so-called schemata whose function is to express grammatical meanings are not "visible" when written, that is Ki'TaB 'book' and uKTub 'write' where KTB is a root, i-a and u-u schemata, are reduced to writing in the same way.

Of cause, it is a comforting idea to believe in the unique grammar for each language, our duty being to discover it. In reality, the situation is much more complicated and the written word/spoken word dichotomy adds a lot to its complexity.

References

- Baudouin de Courtenay J.A. 1912. On the Relation of Russian Writing to the Russian Language. In Baudouin de Courtenay J.A. Selected Works on General Linguistics. Vol. 2, 209-235. Moscow
- Shcherba L.V. 1957. Baudouin de Courtenay and His Contributions to Linguistic Studies. In Shcherba L.V. Selected Papers on the Russian Language, 85-96. Moscow
- Hockett Charles F. 1961. Grammar for the hearers. In Structure of language and its mathematical aspects. In Proceedings of symposia in applied mathematics, vol. 12, 220-236.