|  |
| --- |
| **Annotation of Morphological errors** |
|  |
| *Anastasiia Mikheeva* |
|  |
|  |
|  |
|  |

Abstract

This paper presents results of morphological annotation of Russian Language Corpus consisting of written tasks done by HL (Heritage Language) and FL (Foreign Language) speakers of Russian. Due to complexity of Russian morphological system Second Language learners as well as Heritage learners tend to make errors in using inflectional and derivational morphemes. Our project aims at investigating the nature of these errors and comparing them in order to identify characteristics in acquisition of morphology by these categories of learners.

**Introduction**

Variability of morphemes and the source of morphological errors have been intensely debated in Second Language Acquisition studies. According to the latest findings postpuberty second language (L2) learners often omit or insert wrong affixes as well as inflections more in oral production but less so in written tasks (Andrews, 2001). This paper shows that morphological variability is also characteristic of heritage speakers (or early bilinguals of languages) who were exposed to the family language naturalistically in early childhood but failed to acquire age-appropriate linguistic competence in the language. However, because errors in heritage speakers are more frequent in written than oral production, we took hand-written essays of both categories of Russian language learners to annotate them for morphological errors. At the last, the results of annotation are compared and analyzed in order to identify main tendencies and characteristics in acquisition of morphology by second language learners and heritage speakers of Russian Language.

**Methods and data**

The paper aims at investigating the characteristics of morphological errors made by second language learners and heritage speakers of Russian Language.

For the purposes of morphological annotation, we took the Russian Learner Corpus RULEC that is a collection of texts produced by two categories of Russian language learners: FL or Foreign Learners and HL or Heritage Learners. The original corpus was created by and belongs to Olesya Kisselev and Anna Alsufieva. The Principal Investigators are Professor Olesya Kisselev, Department of World Languages and Literatures, Portland State University (kisselev@pdx.edu) and Professor Anna Alsufieva, Department of World Languages and Literatures, Portland State University ( alsufiev@pdx.edu ).

The materials for RULEC were collected over a period of 4 years from students studying Russian at an American university in a special program that was designed for upper-intermediate and advanced-level Second Language (L2) or Heritage Language (HL) speakers of Russian. The corpus is relatively small: the texts were authored by 36 learners: 17 of the 36 are mainstream American learners who started learning Russian as adults; 19 are heritage speakers of Russian, born in a Russian-speaking country and brought to the US as children or born in the US and raised in a Russian-speaking family. Each author, however, is well-presented in the body of the corpus. RULEC now includes approximately 3,800 written papers ranging from a short paragraph to 8-page research papers (no grammatical or lexical exercises are included in the corpus). The Corpus contains the following metadata: student’s name (pseudonym), gender, language background and language experience of the student (L2 or HL), student’s linguistic level (established through external tests), time stamp (week and academic year when the paper was written), time limit under which the paper was written (timed or non-timed), text type (one paragraph or a long research paper), text function (e.g. narration, argumentation), and whether a paper was written individually or in a group.

For this project available were essays of only 28 students on advanced and intermediate levels. For morphological annotation 132 essays were selected, where 66 of which were written by FL and 66 by HL speakers of Russian.

Each of the essays presents a single .txt file. The project aims at annotation of morphological errors done by Russian Language learners. The annotation was performed manually using UAM Corpus Tool 3 and self-designed annotation model.

Following the linguistic conversation, we posed the following research questions:

● What types of morphological errors are more frequent among intermediate FL and HL learners and therefore present more difficulties?

● How FL and HL learners on intermediate levels differ from each other in terms of morphology acquisition?

**Morphological system of Russian**

The morphemes in Russian are divided into 2 groups: derivational and inflectional. Derivational suffixes, prefixes and postfixes were grouped together into a larger category called affixes.

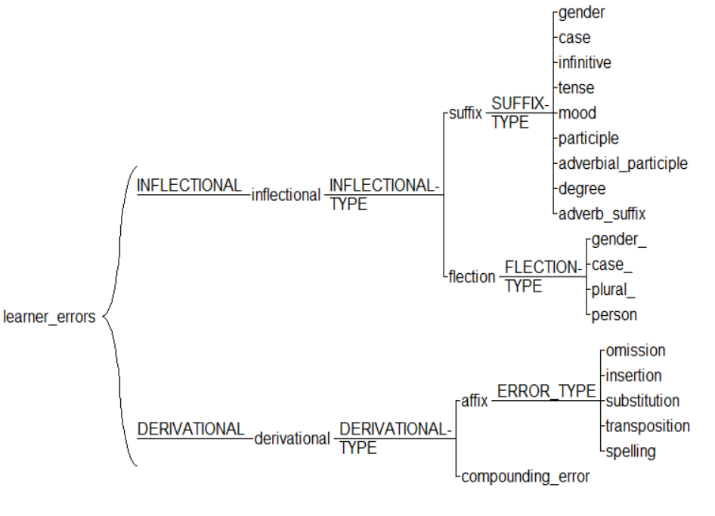
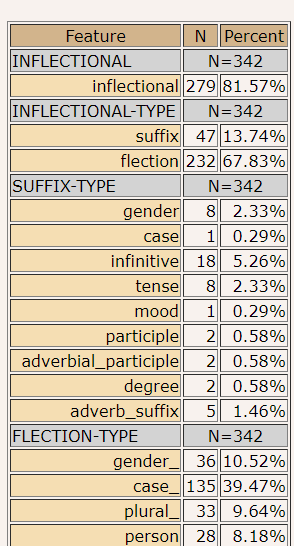
Derivational suffixes in Russian include suffixes that help form words (стол – стол-ик, груз-ить – груз-чик, бел-ый – бел-изн-а). Derivational prefixes are also common in the language to build words (пере-делать, с-делать, до-делать, вы-делать). One of the most common ways to form a word is compounding where two different roots are added in order to make a single word (пар-о-ход, пут-е-провод). Postfixes are suffixes that are attached to a word at the most end after all derivational suffixes (-ся/-сь, -те, -то, -либо, -нибудь).

Inflectional morphemes are divided into suffixes and inflections (or flections). Suffixes refer mainly to the infinitive of a verb (-ти/-тъ), to a stem of a verb in the present or past tense (-л-), imperative form of a verb (и), participles (-ущ-, -ющ-, -ащ-, -ящ-, -вш-, -ш-, -им-, -ом-, -ем-, -нн-, -они-, -енн) and adverbial participles (а, -я, -ши, -вши, -учи, -ючи, -ав), comparatives and superlatives by adjectives and adverbs (ее, -ей, -ейш-, -айш-, -ше, -же), plurals of nouns (-/-, -ее-, -ен-), adverbs (-о, - e).

Inflectional morphemes such as flections refer to the categories that change their forms in the process of declension: (мальчик – мальчика, мальчику, мальчиком). Inflected can be nouns and adjectives.

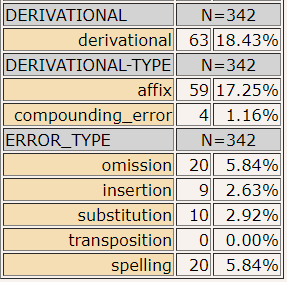
**Annotation Model and specifications**

Because Russian is a highly inflected language the tags involved (see Table 1) are more numerous and elaborate than those typically used for English.

  Table 1. Annotation Model for morphological errors.

A label was assigned to a word that was recognized as having a morphological error. Some words in the corpus presented a difficulty due to the complexity and ambiguity of Russian morphological system.

Derivational errors consist of errors made in using affixes as well as of compounding errors. Types of errors, used for annotating learner corpuses include errors of omission, insertion, substitution, transposition, and spelling. Insertion is when an affix was added by a student and was unnecessary and therefore ungrammatical. Omission is a backward process by which an affix was deleted (omitted) by a student. As for substitution, it is an error that occurs in the process of replacement of a morpheme by another incorrect one when this specific information is forgotten or unknown. Under transposition error we understand an error that is caused by substituting two (or more) sequential affixes. Compounding errors are rare and include the forgotten or omitted vowel (usually o or e in Russian) that are necessary in order to combine two separate word roots together). Tables 2 and 3:



Morphological Analysis

In the process of morphological annotation overall 342 tokens were annotated manually for errors that consisted both of inflectional and derivational morphemes. Inflectional morphemes were identified in 279 cases which amounts to 81.57%, whereas derivational morphemes (see Tables 2 and 3)

were tagged just in 63 cases or in 18.43% of the cases. The following tables (Tables 2 and 3) display what type of inflectional and derivational morphemes had greater numbers.

The role of morphology is important in determining the grammatical functions of phrases. According to the statistics (see Tables 2 and 3) the prominence of inflectional errors divided into suffixes and flections is striking. The largest number of all mistakes have flections.

In Russian each of syntactic categories has its morphological characteristics. Grammatical features of nouns are gender, number, and case. Adjectives agree with nouns in all these categories.

However, they can show the degree feature which is also reflected in their endings. The verb reflects the main grammatical features such tense, aspect, voice, person, gender, number, and case. Russian adverbs show the degree only, whereas pronouns agree with nouns in gender, number, and case.

Regarding the errors in inflectional morphemes the main problem for Russian language learners presented the inflections. In most of the cases students made errors in types of cases. It is known that Russian Language has 6 cases: nominative, genitive, dative, accusative, instrumental, prepositional which encounter significant difficulties in their usage. The declination system is extremely complex which results in high percentage of the mistakes done by all learners. Inflectional suffixes take the second place among all morphological errors. The table 2 illustrates that in 18 cases the infinitive suffixes created more problems than for example suffixes of adverbial participles or adverbs.

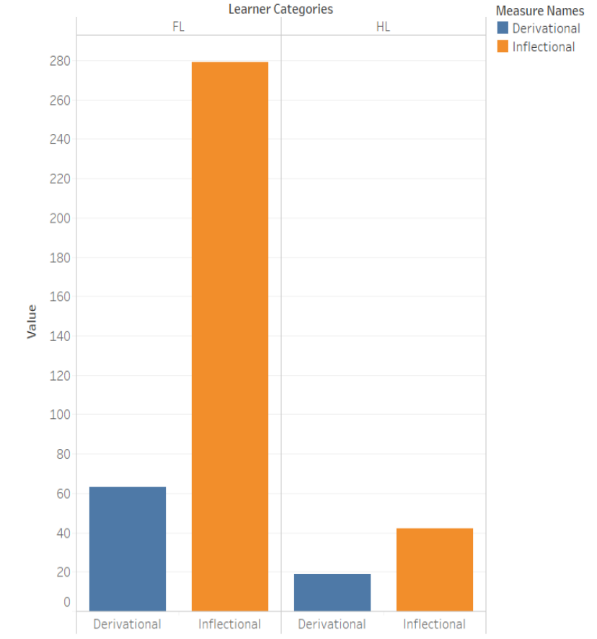
Derivational morphemes consist mainly of affixes of different types, such as: prefixes, suffixes, and postfixes. According to the statistics, the total number of derivational errors is 63 cases. The number of incorrect affixes attached to the words in order to form a syntactic category equals 59.

Derivational compounding errors were divided into 5 most common types which are: omission, insertion, substitution, transposition, and spelling. The most common error type was omission which means that students forgot to attach an affix at all. At the second place was spelling of affixes. In 2.92% cases students substituted a morpheme automatically which was obviously forgotten or unknown. The number of incorrectly inserted derivational morphemes counts to 9. As a result of the first statistics we can conclude that the largest number of all errors have inflectional morphemes. Most of students had significant troubles with determining type of case as well as choosing an appropriate personal verb ending in the process of conjugation.

The next step will be to discover the relationship between types of morphological errors and learner categories (foreign and heritage).

Evaluation

First, let us have a look at the nature of

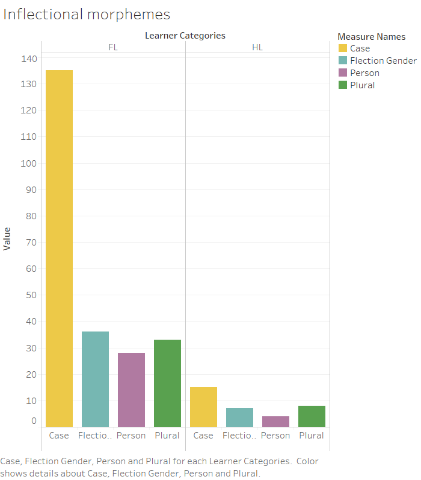
the errors.

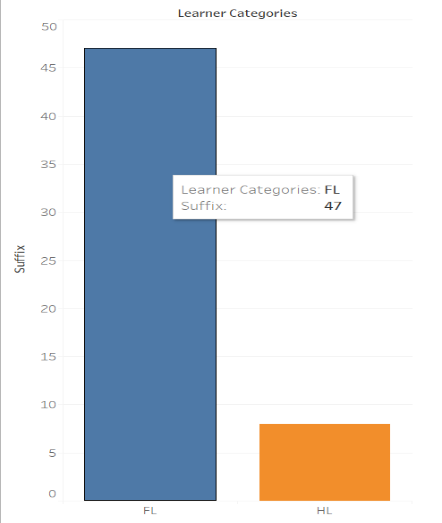
Graph 1.

The following graph 1 illustrates the prevalence of inflectional errors made by both categories of learners. Moreover, as we can see the overall number of all types of morphological errors are significantly higher by foreign language learners than by heritage learners, despite the fact that the number of essays of both categories is equal, as well as topics of essays and proficiency levels are the same. The reasons for such a tendency is obviously the exposure of HL to language in the family since their childhood, whereas foreign learners did not grow up in a natural language environment. Heritage language learners are born into households where a language other than English is spoken and they are bilingual, "at least to some degree" (Valdés 2000: 375). It might seem that this category of speakers will not have many troubles connected with word formation and inflections. It is partly true. Heritage learners all share certain characteristics. First, they are bilingual, and few, if any, bilinguals are equally competent in two languages in all areas. However, their skill levels may vary. Furthermore, the speakers possess two different language systems that may overlap and cause such language problems as interference. Although skill levels vary, heritage learners across languages share a lack of familiarity with the full range of stylistic registers available to the educated native speaker (Zemskaia, 2001).

The results show us that the HL do not have a great number of morphological errors in comparison with foreign students, however we observe some difficulties not only in inflectional morphology, but also in derivational.

Let us have a closer look at the most dominant category: inflectional morphemes (Graph 2). Both types of learners struggled finding a correct case of the words, which is one of the most important characteristics of nouns and adjectives. On the second place we can see the variable flection\_gender. Surprisingly, both HL and FL made more errors in formulating a plural form of words than choosing a correct person.

Graph 2.

Possible reasons for having similar patterns is the complexity of Russian morphology. For FL, the case system is the most difficult one to comprehend because it is not present in the language system of their first language. Especially, instrumental, and prepositional cases are among the hardest to be understood by an L1 English speaker. There is also a lot of confusion with Russian affixes as most of the

Graph 3.

grammatical meanings of words are expressed within the word itself by inflections. One inflection usually expresses several morphological meanings that might cause that confusion even by HL speakers. Also, why plural is among the most common mistakes is that each word (nouns, adjectives, or numerical adjectives) when put in a case, demonstrate different plural endings. Not only for FL but also for HL learners, plurals is one of the most important issue.

As for derivational suffixes (Graph 3) the number of overall cases is with 47 errors by FL students and almost 10 errors more than twice less than by inflections. Russian verb-forming suffixes are less difficult because we can observe a lot of patterns. However, adjective and noun- forming suffixes are still a big issue for many students.

Conclusion

In this project we presented a morphological model for annotation of errors made in inflectional and derivational morphemes of

Russian Language. The results of errors made by Heritage and Foreign Russian learners were compared in evaluation section of the project and showed that the inflectional morphemes were the most important sticking points to both learner categories. However, the overall number of errors made by FL was more than twice greater than by HL students. The received results are of interest to anyone engaged in morphological analysis of Russian. The HL showed not so high numbers of morphological errors, however we proved that these types of errors are still common in written tasks done by HL learners. However, the overall numbers of errors allow us to conclude that as a future step for further linguistic analysis of HL errors would be investigating factors that influence the acquisition of morphology by HL students and comparing them with FL category of learners. In terms of annotation, it would be also appropriate for future analysis to annotate syntactic and semantic errors.

References

*Andrews, D.* 2001.Teaching the Russian heritage learner: Socio- and psycholinguistic perspectives. Slavic andEastern European Journal, 45 (3), 519–530.

*Bley-Vroman, R.* 1989. The logical problem of second language learning. In Gass, S. & Schachter, J. (Eds.), Linguistic perspectives on second language acquisition (pp. 41–68). New York: Cambridge University Press.

*Kuzmenko E.* 2018. Morphological Analysis for Russian: Integration and Comparison of Taggers. National Research University Higher School of Economics Moscow, Russia.

*Montrul S.* 2011. Morphological errors in Spanish Second Language learners and heritage speakers. Cambridge University Press.

*John Goldsmith.* 2001. Unsupervised Learning of the Morphology of a Natural Language. Computational Linguistics, 27(2):153–198.

*Valdés, G.* 2000. Heritage Language Students: Profiles and Possibilities. In J.Peyton, J. Ranard & S. McGinnis (Eds.), Heritage Languages in America: Preserving a national resource (pp. 37-80). McHenry, IL: The Center for Applied Linguistics and Delta Systems.

*Zemskaia, E.* 2001. Язык русского зарубежья: Общие процессы и речевые портреты [Language of the Russian diaspora: General processes and language portraits]. Moscow & Vienna: Wiener Slawistischer Almanach].