Productive verb prefixation patterns

Jaroslava Hlaváčová, Anna Nedoluzhko

Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

Abstract

The paper discusses a set of verbal prefixes which, when added to a verb together with a reflexive morpheme, change the verb's meaning always in the same manner. The prefixes form a sequence according to the degree of intensity with which they modify the verbal action. We present the process of verb intensification in three Slavic languages, namely Czech, Slovak and Russian.

Motto[1]

Na záhrade sa rozťukal ďateľ.
Chvíľu si poťukal na jednom strome, potom si zaťukal aj na ďalších.
Naťukal sa za celý deň dosť.
Ak sa dnes nevyťukal, bude zajtra pokračovať.
Ale ak sa uťukal, tak nebude.

English translation:

A woodpecker started pecking in the garden.
He pecked for a while on one tree, then he did some more pecking on others.
During the whole day he pecked quite enough.
If he has not pecked enough, he will continue tomorrow.
However, if he tired himself out with pecking, he will not.

---

[1]We thank our colleague Jano Hric from MFF UK in Prague for composition of the mini-story in Slovak.

1. Introduction

Unlike classic dictionaries, electronic ones are potentially unrestricted. They can include any number of words of a given language. Nevertheless, electronic dictionaries still cannot contain all the words. There are a lot of reasons why not all the words can be included in any dictionary. For example, there might be a disagreement among language speakers which words should be considered a part of the language and which should not. These are, at the first hand, foreign words, which in modern Slavic languages are mostly borrowed from English. Also, proper names are often problematic, especially the foreign ones. There exists a considerable discussion concerning neologisms, colloquial forms, casual words and so on.

A casual word is often constructed from an already existing word by adding an already existing prefix or suffix. These words are the subject of our investigation. In this paper, we will focus on imperfective verbs in Czech, Slovak and Russian, which have a rich potential of accepting prefixes.

In the three mentioned languages, we have found a set of prefixes that can be potentially added to almost all imperfective verbs. Being added to a verb together with a reflexive morpheme, each of these prefixes represents a modification of the verb original meaning. More specifically, this construction (prefix plus reflexivization) changes the intensity of the action. Some useful remarks on this topic can be found in Isachenko (1960). The Czech prefixes with the intensification meaning are mentioned also in Kopečný (1962).

According to the degree of intensity, these prefixal verbs can be organized in the similar way as the degrees of comparison for adjectives or adverbs. Therefore, although by a certain stretch, such kind of prefixation (plus reflexivization) may be called "verb intensification".

The analysis of verb intensification will be proposed for three Slavic languages: Czech, Slovak and Russian. However, we assume that this derivation pattern might hold for other Slavic languages as well.

2. Intensifying prefixes

Majority of Czech, Slovak and Russian imperfective verbs might be "intensified" according to a simple rule:
1. add a prefix from a special closed set of prefixes and
2. add a reflexive morpheme, namely the reflexive particle *se*, *si* in Czech, *sa*, *si* in Slovak, and the reflexive suffix *-ся* (*-сь*) in Russian.

Prefixes to be discussed are summarized in Table 1, where they are sorted according to an intensity they modify the verbal action expressed by the basic verb. For easier referencing in the following text, we assigned initial letters to these prefixes (in Latin alphabet). If we mention, for instance, the prefix Z-, we will have in mind all the prefixes in the column superscribed Z- in Table 1, namely *za-* in Czech and Slovak and *за-* in Russian.

Abbreviation	R	P	Z	N	V	U	I	D
Czech	roz-/roze-	po-	za-	na-	vy-	u-		
Slovak	roz-/rozo-	po-	za-	na-	vy-	u-		
Russian	раз-/рас-/разо-	по-	за-	на-	вы-	у-	из-/ис-/изо-	до-

Table 1. Prefixes of intensification for Czech, Slovak and Russian

We can see that the row containing the Russian prefixes is extended by two columns. It contains the additional prefixes *из-/ис-* and *до-*, which represent the highest degrees of intensity. Although these prefixes also exist in Czech and Slovak, they do not have the intensified meaning.

Table 2 characterizes briefly the individual prefixes, especially their intensification role in Czech and Slovak, Table 3 presents the same for Russian.

Prefix	Verb	Reflexive morpheme	Meaning
R-	X	*se/sa*	start to X
P-	X	*si**/*se/sa*	X calmly, mostly in pleasant manner
Z-	X	*si**/*se/sa*	X taking some time, enjoying it
N-	X	*se/sa*	X intensively
V-	X	*se/sa*	X intensively, with a satisfaction
U-	X	*se/sa*	X to exhaustion

Table 2. Survey of intensifying Czech and Slovak verbal prefixes. The asterisk (*) in the table means that the morpheme si *is employed for originally non-reflexive verbs. In the case of reflexivum tantum, the morpheme* se/sa *remains.*

Dividing the information into two tables is intentional. We want to point out the sameness of the Czech and Slovak languages, and their differences in comparison to Russian. They concern not only different meanings of some prefixes, but also the reflexivity of resulting verbs, especially of those related

Prefix	Verb	Refl. morpheme	Meaning
R-	X	*-ся*	start to X or strengthen the intensity of X
P-	X	*0/себе*	X during a relatively short time period
Z-	X	*-ся*	immerse into X for a longer time period
N-	X	*-ся*	exhaustively X, often with a positive result
V-	X	*-ся*	X until exhaustion
U-	X	*-ся*	X until exhaustion
I-	X	*-ся*	X until total exhaustion, often with a neg. result
D-	X	*-ся*	X with high intensity leading to a neg. result

Table 3. Survey of intensifying Russian prefixes.

to prefixes P- and Z-. We will demonstrate the differences on examples in the following section.

Although the intensifying meaning of the individual prefixes is always the same, their intensity is not absolute in neither of the languages. The meanings overlap. Only the prefix R- might be assigned an absolutely lowest intensity in Czech and Slovak. In both languages, it represents beginning of an action, thus the weakest intensity. On the contrary, the prefix U- represents the highest intensity.

In Russian, the overlapping seems to affect all the prefixes, including the extreme ones. Here, we might spot another ordering, namely according to a positive or negative result of the verbal action. We present two pictures (1 and 2) demonstrating the sequence of the prefixes according to their (possibly overlapping) intensity (see also similar pictures in Hlaváčová and Nedolužko (2012) and Hlaváčová and Nedoluzhko (2013)). In Russian, the intensity of the last five prefixes is very similar (see the the right end of Figure 2).
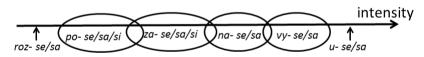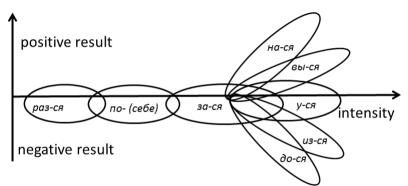

Figure 1. Czech intensifying prefixes

Figure 2. Russian intensifying prefixes

3. Examples

In what follows, we present a set of examples for all studied prefixes in each of languages mentioned above. Every example consists of one sentence in all the languages. The sentences are not authentic, as it would be really very difficult to find the same sentence in three languages, especially for such a rare phenomenon. Even so, for some prefixes it was not possible to translate the same sentence into all three languages by means of verb intensification. The Czech and the Slovak languages are very similar, so translations between them was easy but the meaning of several Russian prefixes is so different that some sentences had to be reformulated. Otherwise the translated sentence would sound unnatural. The prefix Z- varies in Russian so much that it was impossible to use the same prefix for all three translations. In that case we had to come up with a totally different sentence to demonstrate the prefix meaning.

There are authentic examples from the Czech National Corpus SYN (CNK) cited in Hlaváčová (2009). The works Hlaváčová and Nedoluzko (2012) and Hlaváčová and Nedoluzhko (2013) contain more examples in Czech and in Russian. The latter were found in the Russian National Corpus (RNC) and on the Internet.

Prefix R-

Czech	*Před vystoupením se rozezpíval.*
Slovak	*Pred vystúpením sa rozospieval.*
Russian	*Перед выступлением он распелся.*
In English	*To be prepared, he started to sing before the performance.*

Prefix P-
  Czech   *Pohrál si s nastavením parametrů.*
  Slovak   *Pohral si s nastavením parametrov.*
  Russian   *Он немного поиграл с настройкой парамметров.*
  In English   *He played a bit with the parameter settings.*

Prefix Z-
  Czech   *Zatrénoval si v novém dresu, starý už byl roztrhaný.*
  Slovak   *Zatrénoval si v novom drese, starý už bol roztrhaný.*
  In English   *He enjoyed training in a new sports dress, the old one had been torn.*
  Russian   *Он заработался и забыл про обед.*
  In English   *He buried himself to work and forgot to have a lunch.*

Prefix N-
  Czech   *Po vystoupení se naděkoval víc než minule.*
  Slovak   *Po vystúpení sa naďakoval viac než minule.*
  Russian   *После выступления он никак не мог наблагодариться.*
  In English   *After the performance, he aknowledged the applaus more than usually.*

Prefix V-
  Czech   *Punťa se na zahradě vyběhal do sytosti.*
  Slovak   *Dunčo sa na záhrade vybehal do sýtosti.*
  Russian   *Шарик выбегался досыта в саду.*
  In English   *The dog ran in the garden as much as it liked.*

Prefix U-
  Czech   *Na komedii se uchechtal až k slzám.*
  Slovak   *Na komédii sa uchechtal až k slzám.*
  Russian   *На комедии он ухохотался до слез.*
  In English   *On the comedy, he roared with laughter, almost to tears.*

Prefix I-
  Russian   *Этот автор уже совсем исписался.*
  In English   *This author has exhausted all topics and he has nothing to write about.*

Prefix D-
  Russian   *Первоклашка добегался до перелома ноги.*
  In English   *The first-grader ran around so much that he ended up with a broken leg.*

4. Intensified verbs in dictionaries and homonymy

The examples above show that the meaning of the prefixes is quite clear and it changes the meaning of a basic verb always in the same way. Some prefixal verbs have been already adopted in the language, becoming part of the common vocabulary. One of such words is the verb *rozesmát se – rozosmiať sa – рассмеяться* (*to start laughing*).

However, in many cases, the prefixal verbs of such kind are uncommon. They are used only occasionally and that is why they cannot be found in dictionaries. This is the case of most verbs with the basis *ťukať* (*to peck* in English) from the Slovak motto in the beginning of this paper. It is certainly possible to paraphrase this mini-story without using the intensified verbs. However, using this kind of verbs enables to express the story in a better, more colorful and witty way, and this is the reason why such casual words appear, though not very often. For humans, it is easy to understand such verbs, though they do not appear in any dictionary. It is sufficient to understand the prefix, the verb and its reflexivization.

With the means of the verb intensification, there can be created verbs that already exist in a common vocabulary of a language but have a different meaning. Such verbs are homonymous. An example is the Czech verb *usmát se*. In The Dictionary of Contemporary Czech (SSJC), this word is explained as "express gladness by a smile". This meaning differs essentially from the intensified meaning "to be totally exhausted by laughing" formed by prefixation + reflexivization. Compare the following Czech examples:

<u>Usmála se</u> *na svého přítele.*
(*She smiled at her friend.*)

Smála se tak dlouho, až <u>se</u> *téměř* <u>usmála.</u>
(*She laughed for such a long time that she was completely worn out by laughing.*)

Unlike academic dictionaries of Czech and Slovak (e.g., SSJC; SSJ), the Dictionary of Russian (MAS 1999) contains many uncommon intensification meanings. For example, for the verb *расписаться*, the dictionary gives both common meanings (*sign one's name* and *register one's marriage*) and a relatively rare intensifying meaning (*get into a writing vein*). However, it does not include intensification verbs systematically, for example, the saturative meaning of *наплаваться* (*to swim a lot*) is presented, but not *наныряться* (*to dive a lot*).

5. Verb intensification as an inflectional morphological category

The question is how to treat intensified verbs within automatic language processing, namely how to lemmatize them. It is commonly accepted that a lemma has the same prefix as all of the word-forms which can be derived from it. The exception for the Czech and Slovak are the negation prefix *ne-* and superlative intensifying prefixes *nej-/naj-/наи-* of adjectives and adverbs in all the analyzed Slavic languages. This gave us an idea of using the notion "intensification" for regular composing prefixal reflexive verbs with the listed set of prefixes. Taking this into account, it makes sense to use the infinitive of unprefixed verbs as the lemma of the intensified verbs. For example, the Czech intensified verb *vyběhal se* with the prefix V- (see Section 3) will be lemmatized as *běhat* (*to run*).

This decision seems to be more precise than using the lemma with the prefix, because this will emphasize that the prefixal word-form belongs to the paradigm of the basic unprefixed verb. Indeed, the prefix does not change the meaning of the verb in these cases, it only modifies its intensity.

Another advantage of such decision is the possibility to resolve the homonymy of the forms mentioned in 4. In the first example, the word-form *usmála* will be assigned the lemma *usmát (se)* (*to smile*) [2]. In the second case, the lemma will be *smát (se)* (*to laugh*). With such a lemmatization, the information about the degree of intensification can be observed as an inflectional morphological category for verbs. This category can be called intensification, similarly as degrees of comparison for adverbs and adjectives. This morphological category can be realized with the means of the affixes in question together with a reflexive morpheme of a given language.

Thus, we believe that adding one of the listed prefixes and a reflexive morpheme (with the exception of the prefix P- in Russian — see Table 3) to a verb does not create a new verb but a new form of the same verb.

6. Automatic recognition of intensified verbs in texts

Morphological analysis assigns morphological characteristics and lemmas to each word. For this purpose, extensive morphological dictionaries, including all[3] words of the given languages, are used. Such dictionaries exist for

---

[2]The question of the reflexive morpheme being a part of the lemma is not discussed in this paper.

[3]as many as possible

all three languages under analysis. However, real texts still include unknown words. For recognition of such words, so called guessers are usually used. They are able to guess the morphological characteristics of unknown (out of vocabulary, or OOV) words. The guesser should be able to distinguish intensified verbs and to assign them an appropriate set of morphological properties, including a lemma.

First of all, intensified verbs should be recognized in a text. If they are included in the dictionary, they will be analyzed in the same way as other words. In that case, there is no reference to their intensified nature, because dictionaries often do not include this sort of information. This concerns common verbs that have already been lexicalized.

For an unknown word, there is quite easy procedure how to check if a verb in question is the result of the intensification. The following two conditions must be fulfilled:

- The word can be split into two strings AB, where A belongs to the set of intensification prefixes, B is an imperfective verb (recognized as such by the morphological analysis).
- There is a reflexive morpheme either attached to the verb (in case of Russian, with the exception of the prefix P) or within a certain span around the word in the same sentence (in case of Czech and Slovak).

Under these conditions, there is a strong probability that the unknown word might be considered to be an intensified verb. We have carried out a pilot experiment for recognition of intensified verbs in the Czech national corpus SYN2000 (CNK). We extracted all unknown words starting with given prefixes that are immediately followed by the reflexive *se* or *si*. We ignored sentences with the reflexive before the word, as it would be more difficult to distinguish, whether the reflexive belongs really to the word. Then, we removed the prefix and analyzed the rest of the word. If the rest did not appear to be an imperfective verb[4], we did not take it into account. The resulting sets of candidates have been manually investigated and the intensified verbs have been collected. The number of candidates and real intensified verbs for each prefix is presented in Table 4.[5]

---

[4]We also excluded transgressives from the set, as they are very rare, but homonymous with other parts of speech.

[5] For such kind of results, it is common to calculate precision, but due to small amount of data it is not statistically significant.

Prefix	Candidates	Real intensified verbs
roz-	25	20
po-	11	5
za-	6	4
na-	11	2
vy-	2	0
u-	7	2

Table 4. Number of candidates for intensified verbs.

The figures presented in Table 4 show, that the rule for the recognition of intensified verbs is not perfect. However, the false findings were usually due to various errors in the original texts. For example, there are typos such as the word *postekl*, where *stekl* has the lemma *stéci — to flow down*, instead of the correct *posteskl* (*to express dissatisfaction*). Also, words used in a rather unusual form which is missing in the dictionary are problematic, e.g., the verb *vyhlo* as a variant of *vyhnulo* (*avoided*).

Another evidence that could help recognizing intensified verbs in a sentence is the use of certain intensifiers. For example, the typical intensifiers for the prefix U- are *do sytosti, dosyta* (to one's heart's content), *úplně* (completely), *k smrti* (to death), *do (úplného) vyčerpání* (to (complete) exhaustion) in Czech, *dosýta, úplne, k smrti, do (úplného) vyčerpania, do sýtosti* with the same meanings in Slovak, and *совсем* (entirely, totally), *досыта* (to one's heart's content), *до смерти* (to death) in Russian.

The information about the verb intensification can be useful in some applications of automatic language processing, e.g., in machine translation. If the input and the output languages have similar intensification degrees, the same principles for words in both languages can be applied. For example, the Slovak word *rozťukať sa* (*to start pecking*) from the Motto will be translated into Czech as *rozťukat se*, although this verb possibly does not exist in any morphological dictionary of either of these languages. In the languages without intensification, for example in English, the special rules for translating this prefix should be applied. For example, the prefix R- is likely to be translated using the expression *start to*.

In automatic language processing, it is also important to recognize the homonymy between basic and intensified verbs (see the example above with the Czech verb *usmát se — smile* or *to be completely exhausted by laughing*). Such

a distinction seems to be difficult at first sight, but there are certain signs that help revealing intensified verbs in a sentence. One of such signs is the fact that intensified verbs typically lack an object, even in case of transitive verbs. Once the verb is intensified, its object is generalized and cannot be explicitly expressed in the sentence. Cf. *číst knihu* (*read the book*) — *učíst se k smrti* (*to read to death*), but not **učíst se knihu k smrti* (*to read the book to death*). The same can be claimed about most other verbs that include objects or adverbials in their valency frames. The deeper analysis of syntactic and semantic constraints of the intensification pattern in Russian is provided in Khoroshkina and Nedoluzhko (2014).

7. Conclusion

We presented a set of productive word formation patterns composed of a prefix and a reflexive morpheme. Using the examples of Czech, Slovak and Russian, we have shown that this procedure can be applied to almost any imperfective verb in the above mentioned languages. This does not change the basic meaning of the verb, but only the intensity of the action which is expressed by this verb. The intensification resembles the degrees of comparison for adverbs and adjectives. Inspired by this similarity, we have called it "verb intensification".

The presented approach implies treating the verb intensification differently from the current praxis. We propose to lemmatize newly created verbs with an unprefixed infinitive, thus including the form of intensification into the paradigm of the unprefixed verb.

As for semantic characteristics, the frequency of occurrences of the intensification patterns is higher with the verbs that have an animate actor, who is able to control the action or at least to perceive it. Therefore, intensified verbs such as, e.g., *nalyžovat se* (*to ski one's fill*) and *naradovat se* (*enjoy oneself to the fullest*) sound much more natural than *napraskat se* (*to crack one's fill*) or *naslyšet se* (*to hear one's fill*), though they are not completely impossible. The same applies to the semantic feature of iterativeness: the action has to be "repeatable" to be able to take the intensification pattern, cf. *?naumírat se*, which is only compatible with the context of computer games, where dying can be considered a repeatable action. However, the more detailed analysis of semantic constraints of this kind should be provided for all analyzed languages.

Furthermore, the word-formative patterns of different groups of verbs are of different degree of productivity. More than that, the prefixes themselves are

not equally productive in different analyzed languages. One could possibly assume that in Czech and Slovak, the verb intensification might be more common than in Russian. However, this claim should be proved by a more specific analysis of the use of intensified verbs in these languages. Also the acceptability of certain examples is sometimes ambiguous and requires a broader context. A further research should be performed to bring clarity to these issues.

Bibliography

CNK. Czech national corpus, 2012. <http://www.korpus.cz>.

Hlaváčová, Jaroslava. *Formalizace systému české morfologie s ohledem na automatické zpracování českých textů*. PhD thesis, FF UK, Nám. Jana Palacha 2, 116 38 Praha 1, 2009.

Hlaváčová, Jaroslava and Anna Nedoluzhko. Intensifying verb prefix patterns in czech and russian. In *TSD 2013. Proceedings*, volume 8082 of *LNCS*, pages 303–310, Berlin / Heidelberg, 2013. Západočeská univerzita v Plzni, Springer Verlag. ISBN 978-3-642-40584-6.

Hlaváčová, Jaroslava and Anja Nedolužko. Příklad pravidelných slovotvorných vzorců v automatickém zpracování češtiny a ruštiny. In Horváth, Tomáš, editor, *Zborník príspevkov prezentovaných na konferencii Informačné technológie – Aplikácie a Teória, ITAT*, pages 53–56, Košice, Slovakia, 2012. Univerzita Pavla Jozefa Šafárika v Košiciach. ISBN 978-80-971144-1-1.

Isachenko, A.V. *Grammar of Russian language in comparison with the Slovak*, volume V.II. Jazyki slavjanskoj kultury, Bratislava, 1960. Grammaticheskii stroi russkogo yazyka v sopostavlenii s slovackim.

Khoroshkina, Anna and Anna Nedoluzhko. "Vchera nasochinyalsya voroh strok": Productive circumfixal intensifying patterns in Russian. In *Papers from the Annual International Conference Dialogue 2014*, volume 13 (20) of *Computational Linguistics and Intellectual Technologies*, Moscow, 2014.

Kopečný, František. *Slovesný vid v češtině*. Nakladatelství ČSAV, Praha, 1962.

MAS 1999. *Malyj Akademicheskii Slovar russkogo yazyka*. Poligrafresursy, Moscow, 1999. The Small Academic Dictionary of Russian.

RNC. Russian national corpus, 2011. <www.ruscorpora.ru>.

SSJ. *Slovník slovenského jazyka*. Bratislava: Vydavateľstvo SAV, 1959–1968.

SSJC. *Slovník spisovného jazyka českého*. Praha, Academia, 1989.

Address for correspondence:
Jaroslava Hlaváčová
`hlavacova@ufal.mff.cuni.cz`
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics, Charles University in Prague
Malostranské náměstí 25, 118 00 Praha 1, Czech Republic