Intensifying Verb Prefix Patterns
in Czech and Russian⋆

Jaroslava Hlaváčová and Anna Nedoluzhko

Charles University in Prague
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
{hlava,nedoluzko}@ufal.mff.cuni.cz

Abstract. In Czech and in Russian there is a set of prefixes changing the meaning of imperfective verbs always in the same manner. The change often (in Czech always) demands adding a reflexive morpheme. This feature can be used for automatic recognition of words, without the need to store them in morphological dictionaries.

1 Introduction

An automatic text processing usually starts with a morphological analysis trying to recognize all the words and to assign morphological features to them. In practice, morphological dictionaries are used for that purpose. However, regardless of dictionary size, we meet words which are not included in the dictionary. There are several kinds of such words. So called guessers are used for their recognition. In the present paper, we will examine a special group of verbal prefixes that change imperfective verbs always in the same manner. This feature allows an easy and reliable recognition and subsequent analysis of such verbs. We will present our results on two representatives of Slavic languages — Czech and Russian.

2 Verb Intensification in Czech and Russian

The majority of Czech and Russian imperfective verbs have an interesting ability to be associated with one of several prefixes that change the meaning of the verb in a way that might be called intensification. With one exception on the Russian side, those prefixal verbs must be modified with a reflexive morpheme *se/si* in Czech and *-ся/-сь* in Russian. The prefixes are: *roz-, po-, za-, na-, vy-* and *u-* for Czech and *раз-(рас-), по-, за-, на-, вы-* and *у-* for Russian, the set could be extended with the prefix *из-* for Russian. This prefix has also a Czech equivalent, namely *z-*, but not in the context we are going to present now.

The order we introduced the prefixes is not accidental. With a certain tolerance, both sets can be viewed as sequences ordered according to an intensity of the process

---

described by the verb attached. This is the reason why we call this type of prefixation a **"verb intensification"**.

The meaning of the prefixes could be derived from short stories in both languages. These stories are not authentic, they were compiled by the authors. The English translations are not very good because it is really hard to translate the prefixal verbs of such kind into English properly. In majority of cases, the pure verb would be a better translation, but we wanted to stress that the meaning of the sentence lies primarily in the prefixal verb.

The Czech story with the intensification of the verb *lyžovat* (*to ski*):

Czech *V zimě jsem jela do Itálie* **lyžovat**.
English *(In winter, I went to* **ski** *to Italy.)*
Czech *Marie* **si** *chtěla se mnou* **zalyžovat**.
English *(Marie wanted to* **enjoy skiing** *with me.)*
Czech **Rozlyžovaly** *jsme* **se** *poměrně rychle.*
English *(We* **got into skiing** *fairly quickly.)*
Czech *První den jsme* **si** *jen tak zvolna* **polyžovaly**.
English *(The first day we have just* **skied a little**.)*
Czech *Další dny jsme* **se** *už* **nalyžovaly** *víc.*
English *(Other days we've* **skied more**.)*
Czech *Poslední den jsem už měla pocit, že jsme* **se vylyžovaly** *hodně.*
English *(The last day I have felt that we have been* **skiing lot**.)*
Czech *Už jsme musely odpočívat, abychom* **se neulyžovaly**.
English *(We had to rest no to be completely* **exhausted by skiing**.)*

The Russian story with the intensification of the verb *плавать* (*to swim*):

Russian *Этим летом я ездил в Италию на месяц* **плавать**.
English *(In summer, I went to Italy to* **swim** *for a month.)*
Russian *Первая неделя — это только* **расплаваться, поплавать себе** *в своё удовольствие.*
English *(The first week is good just to* **get into swimming, to enjoy the swimming**.)*
Russian *Однажды я так* **заплавался**, *что даже пообедать забыл.*
English *(Once I* **forgot myself in swimming** *so much, that I missed my dinner.)*
Russian *К концу третьей недели я наконец* **наплавался**, *а к середине четвёртой уже* **уплавался** *до изнеможения.*
English *(By the end of the third week I've* **swimmed to my heart's content** *and by the middle of the forths week I was* **completely tired of swimming**.)*
Russian *В конце месяца у меня было такое ощущение, что я уже весь* **изплавался**, *просто* **выплавался** *напрочь и море просто видеть не могу.*
English *(By the end of the month, I felt* **totally exhausted by swimming, swimming has completely drained my spirit** *and the sea already ticked me off.)*

The figures 1 and 2 illustrate the distribution of meanings of the prefixal verbs with intensification meanings in both languages.
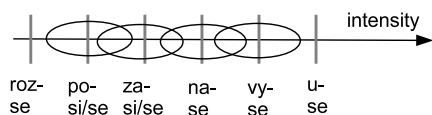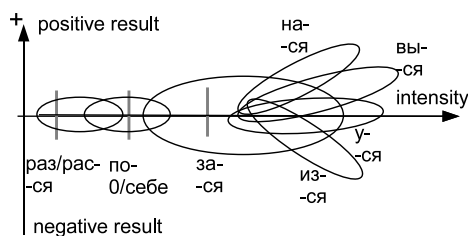
Fig. 1. Czech intensifying prefixes

Fig. 2. Russian intensifying prefixes

In the Czech case, there are two strict points at both ends of the scale. They belong to the prefixes *roz-* and *u-*. The former one represents a starting stage of a process, while the latter one expresses its final stage. Other prefixes cannot be placed on the axis so unambiguously, which is indicated by the overlapping elipses.

In Russian, the situation is similar but not the same. The first three prefixes correspond with their counterparts in Czech, though the prefix *за-* stretches beyond common limits. Also the prefix *раз-(рас-)* is not placed so strictly as in Czech, it is more vague. The last four prefixes (*вы-, на-, у-, из-*) not only overlap in their intensities, but have another dimension of meaning. Although their intensity is roughly the same, their semantic and pragmatic properties differ. This fact is schematically described by the perpendicular axis expressing a scale of positivness / negativness of the process. However, the meanings of individual prefixes are also vague and overlapping. The picture shows the meanigs of prefixes only schematically. There is a more detailed description of the mentioned prefixes and their connection with reflexive morphemes in [1].

The following section presents a complete overview of the prefixes, their meanings and connection with reflexive morphemes. Every pair of prefixes (for Czech and for Russian) is represented by a word-forming pattern. The letter X stands for an imperfective verb. Adding the given prefix and reflexive morpheme, we derive the appropriate intensified verb. We always explain the meaning and, if needed, also a difference between the two languages. The meaning explanation is usually quite vague. A better understanding is usually visible from examples that accompany every prefix in both languages. The Czech examples come from the corpus SYN [2], the Russian ones were found in the Russian National Corpus [3] and on the internet.

2.1 roz-X se, раз-X-ся

Meaning: to start X, in Russian also possibly to amplify the intensity.

Examples:

> *Z jara se člověk nejdříve musí **rozlétat**.*
> *(In spring, one must first **get used to flying**.)*
> *Просто **расслушаться** надо, тогда почувствуешь.*
> *(You should just **begin to listen** carefully and then you will feel it.)*

2.2   po-X se/si, по-X 0/себе

Meaning: to X during a certain time period
    Czech + usually rather calmly, enjoy X-ing
    Russian + usually relatively short time
Note: The difference *0/себе* is rather stylistic — the reflexive morpheme is not obligatory, it has an informal tincture.
Examples:

> *Pak vyslechl Aliciny kritické poznámky, pár minut **si popřemítal**, načež prohlásil, že všecko souhlasí.*
>
> *(Then, he listen to Alice's critical remarks, he **was reflecting** several minutes and then declared that all agree.)*
>
> *Анатоль Максимович укрепился на своих ногах, еще немного **пообду-мывал** ситуацию, а затем, очень внезапно, начал проявлять повышенную активность.*
>
> *(Anatol Maximovič fixed himself on his legs, **spent some time on considering** the situation and then suddenly began to behave very actively.)*

2.3   za-X se/si, за-X-ся

Meaning: to X during a longer period
    Czech + enjoy X-ing
    Russian + longer than usually, or even too long
Examples:

> *Loupil tady i loupežník Klempera a **zaloupežil si** tu i slavný lupič Babinský.*
> *(The bandit Klempera robbed here and also the famous robber Babinsky **liked to rob** here.)*
> *Избавляемся от вещичек, которые давно **зависелись** в нашем шкафчи-ке!*
> *(We are getting rid of old stuff, that **is (lit. is hanging) too long** in our cupboards!)*

2.4   na-X se, на-X-ся

Meaning: to X a lot
    Russian + completely, often with a slightly positive result
Examples:

> *Přestože profesor napsal tolik dopisů ... a **napředstavoval se** v desítkách kanceláří, jeho úsilí se bohužel nesetkalo s úspěchem.*
> *(Although the professor wrote so many letters ... and **many times presented himself** in dozens of offices, his efforts unfortunately have not been successful.)*

*Она руку привязала, чтобы не позвонить Новикову и не спросить, какими же это чудесными именами он Евгению Гордееву **напредставлялся**.*

*(She has tied up her hand, in order not to call Novikov and to ask him, which wonderful names he **used (lit. was using, i.e. many times) to introduce himself** to Eugene Gordeev.)*

2.5   vy-X se, вы-X-ся

Meaning: very much X
    Czech + with more or less positive result
    Russian + until exhaustion
Examples:
    *Francis Kennedy se tam uchýlil, aby **se vytruchlil** ze smrti své ženy.*
    *(Francis Kennedy took refuge there to **have enough time to grieve** over the death of his wife.)*
    *Я много **выстрадался** с тех пор, как расстался с вами в Петербурге.*
    *(I've **suffered really much** from the time, when we parted in St.Petersburg.)*

2.6   u-X se, у-X-ся

Meaning: until exhaustion
Examples:
    *Zemřel na těžkou bronchitidu. Prostě **se ukašlal**!*
    *(He died of severe bronchitis. He simply **died of coughing**.)*
    *Всем смешно, а я уже весь **укашлялся и усопливился**…*
    *(Everybody finds it funny, but I'm **completely exhausted by coughing and having the sniffles**.)*

2.7   —, из- X-ся (Only Russian)

Meaning: great intensity, with more or less negative result
Example:
    *За 40 с лишним лет рисования он не **изрисовался**.*
    *(For more than 40 years, he **hasn't weared himself out with painting**.)*

3   Intensified Verbs in the Context of a Common Vocabulary

The previous overview and all the examples show that the situation in Russian and Czech is very similar. The languages differ in productivity, but mostly in shades of meanings of prefixes investigated. That will be the subject of further, more linguistically oriented research.

    The described formation of new meanings is quite rare and is used only in cases where an author wants to stress a specific intensity of the basic word. They are created only occasionally, like the majority of examples presented so far. However, some words

that were formed according to the method described above, are a normal part of the language. Some of them have exactly the same meaning that was described in the previous section, for instance the word *rozesmát se* (in Czech) and *рассмеяться* (in Russian) with the meaning *to start laughing*.

Others have a different meaning. The verb compound *usmát se* is an example. Its common meaning is described in [4] as "to express joy with a smile", as for instance in *Hezky se na mě usmál. (He smiled at me nicely.)*. In our intensified usage, it does not mean a smile, but a laugh, namely its very high, almost exhausting intensity. If we remove the prefix *-u*, we get the verb *smát se (to laugh)*. The following example is not authentic, as it is very difficult to look for a context where the word has different (moreover, very rare) meaning.

Czech: *Celý večer jsme se něčemu smáli, až jsme **se** skoro **usmáli**.*

English: *(We have been laughing all the evening, until we **almost died from laugh**.)*

4   Use in Automatic Language Processing

4.1   Morphological Dictionaries and Guessers

At the beginning of any automatic language processing task, one should detect the words in the text. For this purpose, the morphological analysis is used, which assigns each word its morphological characteristics. These are e.g. lemma, part of speech and according to the part of speech some other morphological values, such as grammatical gender, number, case, time and so on. These values are usually expressed by means of morphological tags. The lemma for verbs in Czech and Russian (but also in many other flective languages) is infinitive.

Today, many languages have a morphological dictionary that should contain all commonly used words. These dictionaries are used for word detection and analysis. Words that are not included in the dictionaries must be detected and tagged according to some heuristics. These heuristic tools are called guessers. Commonly, a guesser may use any possible rule observed in different texts of the language in question. The observation may be either "manual" (rule-based guessers) or automatic (statistical guessers). Our finding about verb intensification may contribute to better word recognition and analysis. The observation described in the previous sections is considered to be rather a rule-based approach. Due to the high regularity of intensified verbs, it is neither necessary nor possible, to include all of them into a dictionary. In most cases, frequently used intensified verbs are already present in the dictionary (see, for example, the verb *rozesmát se* (*begin to laugh*) described in the previous section); the rest can be detected very reliably.

4.2   Homonymy of Basic and Intensified Verbs

In cases of homonymy between basic and intensified verbs (see e.g. the example above with the Czech verb *usmát se* (*smile*) or (*to be completely exhausted by laughing*)), the most correct approach should be to analyze both meanings. However, due to the low frequency of such cases, we decided not to do so.

Looking at the examples, we have found some factors which can be useful for recognition of different meanings of homonymous verbs. First of all, in the sentences with intensified verbs, verb modifications are generally missing. In the example above, the Czech verb *usmát se* in the sense of smiling is commonly used with the object (*usmát se na něco/na někoho* (*smile at somebody/something*)). If used as an intensified verb, the object is scarcely expressed. This is true for the most verbs which have complements in their basic meaning (either objects or adverbials).

Another sign that could help recognizing intensified verbs in the sentence, is use of certain intensifiers. These intensifiers for Czech are, for example, *do sytosti, dosyta* (*to one's heart's content*), *úplně* (*completely*), *k smrti* (*to death*), *do (úplného) vyčerpání* (*to (complete) exhaustion*) and so on, for Russian there are *совсем* (*entirely, totally*), *досыта* (*to one's heart's content*), *до смерти* (*to death*) and so on.

4.3 Lemma of Intensified Verbs

At first glance, the lemmatization doesn't seem to be problematic. By adding a prefix, a new word is created. It seems to be reasonable, to assign this word a new infinitive with the prefix, in the same way as it is done for all prefixal verbs. In linguistics, this kind of derivation is considered to be a case of word formation. However, due to the high regularity of this phenomenon, the intensified verbs may be also observed as a result of a morphological derivation. All prefixes in this case are simply modifying the value of a basic verb, as shown above. For this reason, we consider it to be more natural, to lemmatize these verbs without the prefix. Thus, for example, the Czech verb *rozsténat se* (*begin to moan*) will have the lemma *sténat* (*moan*), not *rozsténat se* (*begin to moan*). The same is true for Russian.

In addition, it would significantly facilitate the distinction between homonymous verbs, examples of which we mentioned in the previous sections. For the homonymy of the Czech verb *usmát se* (*smile* and *to be completely exhausted by laughing*), two lemmas will be assigned: *smát se* for an intensified meaning and *usmát se* for *smile*.

Such lemmatization is also easier to use in further processes of automatic text processing. For example, in machine translation, it is very unlikely that an intensified verb occurs in a translation dictionary, because, as we pointed out above, these words are mostly casual. According to prefixes, the rules for translation would be created. For example, the Czech prefix *roz-* could be translated to English as the construction *start / begin to* + lemma. If we apply this rule to the verb *rozsténat se* (*begin to moan*), the dictionary will translate its lemma as *sténat* (*moan*) and the correct translation into English will be *start / begin to moan*:

Czech: *James Stidham, který to vše sledoval z nosítek,* se *tiše **rozsténal** hrůzou.*
Russian: (*James Stidham, who watched all this from a stretcher, **began to moan** quietly with horror.*)

Similar rules can be applied for all other prefixes which are analyzed here.

4.4 Some Remarks to the Morphological Analysis of Intensified Verbs

Although the evidence concerning the verb intensification presented here is being now just under investigation, a prefix guesser is already being used in the Czech morphology,

namely in the Morfo project [5]. However, the lemmatization in the project has not been implemented as described in the previous sections yet. Verbs, as well as all other words, are lemmatized together with their prefixes.

The recognition procedure of intensified verbs would be useful to apply on all verbs with the presented prefixes, so that in cases of homonymy it could be detected which kind of verb is used and thus which lemma (with or without prefix) should be assigned to it. However, as intensified verbs are not frequent in texts, it seems to be sufficient (at least for the time being), to analyze only the verbs which are not recognized by a morphological analyzer using a morphological dictionary. Due to low frequency of intensified verbs in texts, the risk of an error is relatively small.

The low frequency of the cases in question does not allow us to provide any statistics of how the success rate of morphological analysis improves; the results would be negligible. However, the more precise description of word-formative and morphological characteristics will always contribute to a more precise language processing.

5   Conclusion

In this paper, we presented productive word-formative patterns consisting of a prefix (*roz-, po-, za-, na-, vy-, u-*, in Russian also *iz-*) and the reflexive morpheme. Using the examples from Czech and Russian, we have showed that being applied to almost any imperfective verb, these patterns always change its meaning in the similar way. For the time being, this regularity should be regarded as just an observation. To effectively use this property in NLP, it is necessary to analyse the formal grammar and semantic characteristics of these structures in more detail.

In further research, we also plan to look at other (mainly Slavic) languages, where we also assume the existence of intensified verbs.

References

1. Evgen'eva, A.P.: Malyj akademičeskij slovar'. Institut lingvističeskich issledovanij, Moscow (1999); Slovar' russkogo jazyka v 4-ch tomach
2. Czech national corpus (2012), http://www.korpus.cz
3. Russian national corpus (2011), http://www.ruscorpora.ru
4. Havránek: Slovník spisovného jazyka českého. Praha, Academia (1989)
5. Hlaváčová, J., Kolovratník, D.: Morfologie češtiny znovu a lépe. In: Vojtáš, P. (ed.) Informačné Technológie Aplikácie a Teória.Zborník Príspevkov, ITAT 2008, Seňa, Slovakia, Univerzita Pavla Jozefa Šafárika, PONT s.r.o., pp. 43–47 (2008)