

Computational Linguistics and Intellectual Technologies:  
Proceedings of the International Conference “Dialogue 2016”

Moscow, June 1–4, 2016

## POSSESSIVES IN PARALLEL ENGLISH-CZECH-RUSSIAN TEXTS

**Nedoluzhko A.** (nedoluzko@ufal.mff.cuni.cz)

Charles University in Prague, Prague, Czech Republic

**Schwarz (Khoroshkina) A.** (annakhor@gmail.com)

ABBY; Moscow State University, Moscow, Russia

**Novák M.** (mnovak@ufal.mff.cuni.cz)

Charles University in Prague, Prague, Czech Republic

We present a corpus-based analysis of the use of possessive and reflexive possessive pronouns in a newly created English-Czech-Russian parallel corpus (PCEDT-R). Automatic word-alignment was applied to the texts, which were subsequently manually corrected. In the word-aligned data, we have manually annotated all correspondences of possessive and possessive reflexive pronouns from the perspective of each analysed language. The collected statistics and the analysis of the annotated data allowed us to formulate assumptions about language differences. Our data confirm the relative frequency of possessive pronouns in English as compared to Czech and Russian, and we explain it by the category of definiteness in English. To confirm some of our hypotheses, we used other corpora and questionnaires. We compared the translated texts in Czech and Russian from our corpus to the original texts from other corpora, in order to find out to what degree the translation factor might influence the frequency of possessives.

**Key words:** possessive pronouns, possessive reflexive pronoun, coreference, comparative analysis, parallel corpus, English, Czech, Russian

# ПРИТЯЖАТЕЛЬНЫЕ МЕСТОИМЕНΙΑ В ПАРАЛЛЕЛЬНЫХ АНГЛО-ЧЕШСКО-РУССКИХ ТЕКСТАХ

**Недолужко А.** (nedoluzko@ufal.mff.cuni.cz)

Карлов университет в Праге, Прага, Чехия

**Шварц (Хорошкина) А.** (annakhor@gmail.com)

АВВУУ; МГУ, Москва, Россия

**Новак М.** (mnovak@ufal.mff.cuni.cz)

Карлов университет в Праге, Прага, Чехия

**Ключевые слова:** притяжательные местоимения, возвратно-притяжательное местоимение, кореферентность, сравнительный анализ, параллельный корпус, английский язык, чешский язык, русский язык

## 1. Introduction

Possessivity and how it is expressed by means of possessive and reflexive possessive pronouns is a popular topic in theoretical linguistics. This subject is discussed from many different perspectives, e.g. typological, historical, semantic, syntactic and discursive. The approach used in this paper is contrastive, multilingual and corpus-based.

While analysing the correspondences between coreferential expressions in parallel English-Czech texts (Novák—Nedoluzhko, 2015), we have observed that possessive pronouns in English appear more frequently than they do in the same positions in Czech. This led us to carry out a systematic analysis of such cases. While analysing the Czech counterparts we discovered many language-specific details in Czech that required a comparison to another Slavic language in order to properly interpret this more thoroughly. And we chose Russian<sup>1</sup> for this analysis. The use of possessives has been analysed in detail for each language separately (see Section 2). However, Czech and Russian are typologically and genetically very close, and the rules for the use of possessives are quite similar. So, the general rule both for Czech and Russian is that a reflexive possessive can substitute personal possessive pronouns if it is coreferential with the subject. On the other hand, the use of possessives is not completely identical. For example, the distributive meaning of the reflexive possessive *svoj* is very common in Russian, while in Czech it is completely marginal, see Example 1:

- (1) RU: *U každogo učnogho jest' svoja biblioteka*—CZ: *Každý vědec má \*svou/vlastní knihovnu* [EN: Each scientist has his own library.]

---

<sup>1</sup> One of the reasons is that there is extensive literature on this topic in Russian.

The main concern of this paper is to investigate the use of possessive and reflexive possessive pronouns in English, Czech and Russian. For our analysis, we used a newly created three-language Czech-English-Russian parallel corpus (PCEDT-R, see Section 3), which we provided with automatic word alignment, its manual correction and annotation. For the aligned annotated data, we compiled the statistics of correspondences of the pronouns under analysis in the three languages (Section 4). The next step was to analyse the resulting correspondences. In Section 5, we provide proof for the assumption that Russian possessive and reflexive possessive pronouns occur less frequently than in English, however we also show that they are used significantly more frequently than in Czech. The differences between Czech and Russian are addressed in Section 6. The statistics of correspondences has been provided for the PCEDT-R corpus, the Czech and Russian parts consist of translated texts. The fact that these are translated texts could influence the frequency of pronouns of different types. For this reason, we compared our findings to the original texts (Section 7). The comparison also revealed some interesting differences in the use of possessives in Czech and Russian.

Overall in this paper, we use existing theoretical knowledge from non-corpus-based monolingual research, the annotation of corpus data and our language intuition to formulate hypotheses about the use of possessives in Czech, English and Russian and support them with the statistics from the three-language parallel corpus analysis. We have not found the theoretical basis for all of our assumptions yet, which is planned for the next stages of the research. We believe that the research in this field is helpful both for the improvement of machine translation work (e.g. it would be easier to identify which English possessives should be translated into Czech and Russian and which should not) and for theoretical comparative language analysis.

## 2. Related work

There is a variety of means to express the notion of possessivity (see e.g. a detailed survey in Brykina 2009). In this paper, we are interested, above all, in pronouns. In English, there is one group of possessive pronouns (*his, her, its, their*)<sup>2</sup>, which are translated into Czech and Russian as possessive (*jeho, její, jejich* in Czech, *jego, jeje, ich* in Russian) and possessive reflexive (*svůj* in Czech and *svoy* in Russian) pronouns.

For Czech, the description of personal and possessive reflexive pronouns begins with Svoboda (1880) and is further addressed in a number of theoretical studies and grammars (see e.g. in Gebauer 1890, Trávníček 1951, Daneš—Hausenblas 1962, Paněvová 1986, Dočekal 2000, etc.). The concurrence of possessive and reflexive possessives is described in most detail for syntactic constructions with one predication (*Já čtu svou/?mou knihu = I'm reading my book*) and for cases with an explicitly embedded predication (*Já slyším tě zpívati svou/mou/tvou oblíbenou píseň. = I hear*

---

<sup>2</sup> Here and later, we only speak about possessives in the third person.

*you sing your/my favorite song*<sup>3</sup>). Many examples of different types are systematically analysed, but the results are formulated rather as tendencies than as strong rules and are based on syntactic (Trávníček 1951, Daneš—Hausenblas 1962, Dočekal 2000), semantic (Panevová 1986) or discursive (Čmejrková 1998) criteria. For Russian, the concurrence of possessive and reflexive possessive pronouns is addressed in most detail in Padučeva (1985); the author provides ten distinctions between the different types and explains them with syntactic, semantic and referential arguments. Some non-typical types of control of possessive reflexive were addressed in Fed'ko (2007).

As for the meanings of the reflexive possessive in Russian, Padučeva (1983) considers 6 different types of *svoj*. The basic type is the possessive form of a reflexive pronoun *sebjá* (*oneself*), the same as in Czech; other meanings are derived from the basic one with additional functions added, such as the distributive meaning, the meaning of 'special', 'appropriate', the contrast 'one's own' ↔ 'somebody's else' and so on. In contrast to Czech (Panevová, 1986), Padučeva describes the tendency of *svoj* to take part in different kinds of idiomatic expressions. And in addition to the six meanings of *svoj* introduced in Padučeva (1983), Brykina (2009) mentions that it may be used in sentences where it is semantically redundant, though it also possesses discursive or syntactic functions, such as indicating the focus of attention and maintaining referential connectivity (Brykina, 2009:135ff.).

The study of meaning and distribution of possessive and reflexive possessive pronouns in our analysis is closely connected to the subject of external possessivity. In this respect, we took into account the research provided in Brykina (2005, 2009) and Kibrik (2003) for Russian and Krivan (2007) for Czech.

As for existing parallel corpora including all the languages under analysis, there are the Intercorp (a part of the Czech National Corpus)<sup>4</sup> and the ParaSol<sup>5</sup> multi-language corpora. In Intercorp, parallel data can be excerpted for pairs of languages, namely separately for e.g. English-Czech and Czech-Russian. Texts in both corpora are automatically sentence-aligned, there is no word-alignment. As far as we know, parallel language data have not been used for this kind of analysis yet. The research addressing semantic and pragmatic functions of possessives in Czech mostly relies on the linguistic intuition of the authors and the analysis of constructed or specially found examples.

### 3. Data and methods

Our core analysis is based on the newly created three-language parallel English-Czech-Russian corpus. The English-Czech part of it was taken from the Prague Czech-English Dependency Treebank (PCEDT, Hajic et al. 2012) and translated into Russian (in what follows, the abbreviation **PCEDT-R** will be used for this three-language

---

<sup>3</sup> This Gebauer's (1890) example was used by many researchers addressing this topic.

<sup>4</sup> Texts in Intercorp are taken from the Project Syndicate website (<http://www.project-syndicate.org>).

<sup>5</sup> The project of the Humboldt University of Berlin, <http://www.slavist.de/>

parallel corpus). Given the size of PCEDT, the translation and the manual annotation of word alignment of the entire PCEDT would be extremely time-demanding. We therefore limited the dataset to only 1078 sentences located in the first half of the PCEDT section 19, i.e. the 50 documents from *wsj\_1900* to *wsj\_1949*.

The English part consists of the Wall Street Journal (WSJ) section of the Penn Treebank (Marcus et al., 1999). The Czech and Russian parts were manually translated from the English source sentence by sentence.<sup>6</sup> The linguistic annotation in the English-Czech part of PCEDT-R is provided within the following annotation layers: the lowermost “word” layer (w-layer) representing the tokenized plain text, the morphological layer (m-layer) containing automatic part-of-speech tagging and lemmatization, the analytical layer (a-layer) representing surface dependency syntax, and the deep syntax or tectogrammatical layer (t-layer). The t-layer includes semantic labeling of content words (nouns, adjectives, adverbs, and verbs) and coordinating conjunctions, ellipsis reconstruction, coreference annotation, and argument structure description based on a valency lexicon. PCEDT-R is the same excerpt of texts that was used for analysis of coreferential expressions in English and Czech (Novák—Nedoluzhko, 2015), thus it already contains manual annotation of word alignment for English personal and possessive pronouns and Czech relative pronouns. The Russian part was automatically aligned with the Czech part of PCEDT using GIZA++ (Och and Ney, 2000), which was run on a large amount of parallel Czech-Russian data. The resulting triples containing possessive units (in at least one of the languages used) have been manually annotated and analysed from the perspective of each language separately.

Table 1 shows some of the basic statistics and information related to the present work calculated on PCEDT-R.

**Table 1.** Statistics for PCEDT-R

|  | English | Czech  | Russian      |
|--|---------|--------|--------------|
| texts                                  |         |        | 50           |
| sentences                              |         |        | 1,078        |
| tokens                                 | 26,560  | 25,477 | 25,396       |
| possessive                             | 238     | 106    | 120          |
| possessive-reflexive pronouns          | —       | 91     | 85           |
| morphological and syntactic annotation | yes     | yes    | no (planned) |

<sup>6</sup> The Czech translation has been created for the whole PCEDT (ca. 50,000 sentences, see Hajič et al., 2012) by professional translators. The translation into Russian has been completed recently by A. Schwarz. The Czech and Russian translators were instructed to keep the sentence structure of the source texts. The aim of the research was completely unknown to Czech translators. As for the Russian translator, she was informed that the texts were intended for a comparative study of coreference chains. This fact could affect the translation, therefore we compared the differences between translated texts in Russian and Czech. The results showing the smaller difference between original and translated texts for Russian than for Czech in our corpus are presented in Section 7 below.

To confirm our hypotheses formulated on the basis of examples and statistics from PCEDT and its Russian translation, we also used the examples from *InterCorp*. However, parallel texts there only have sentence automatic alignment and automatic morphological tagging, they are not word-aligned. It means that searching for possessive pronouns there leads to numerous false results. In this study, we do not use the statistics obtained from this corpus. We used it to search for specific triples of sentences (ENG-CS-RU) for cases that we considered to be of interest. In general, over 600 English-Czech-Russian triples have been concerned in *InterCorp*, and only 70 of them are relevant for our analysis.

Our core analysis is based on corpus data, but to confirm some of our general estimates, we also used two short questionnaires for native speakers of Czech. Mainly, they contain questions aiming to find systematic differences between Czech and Russian possessives and are based on the functional description of reflexive possessive pronoun *svoj* in Russian completed by E.V. Padučeva (Padučeva, 1983).

## 4. Statistics

For PCEDT-R, we have calculated the number of occurrences of counterparts of aligned possessive and reflexive possessive pronouns. The calculation has been completed for each of the analysed languages.

Tables 2, 3 and 4 show the statistics for the aligned counterparts for English, Czech and Russian respectively. Possessives in language A may be mapped on the following categories in language B:

- possessive pronouns (the *poss* label, e.g. EN: *his*—RU: *jego*—CZ: *jeho*);
- reflexive-possessive pronouns (the *refl-poss* label, e.g. EN: *his*—RU: *svoj*—CZ: *svůj*);
- nouns, anaphoric zeros, demonstrative and personal pronouns (the *NP* label, see Example 2);

(2) EN: <His> [Steppenwolf's—AN et al.] *board members alone have pledged \$800,000.*—RU: *Только члены правления <Степпенвульф> обещали \$800 000.*

- external possessive expressions, the definite article (in English) or relative clauses (the *other* label, see Example 3, where the possessive meaning is expressed in Czech with external dative reflexive *si*, and Example 8 in Section 5 below for the definite article);

(3) EN: *Glenn and Sharon Beebe of Cincinnati had sued the company in 1981 after installing Burlington carpets in <their> office.*—CZ: *Společnost zažalovali Glenn a Sharon Beebeovi z Cincinnati v roce 1981 poté, co <si> koberce Burlington položili do kanceláře.*

Possessive and reflexive possessive pronouns may remain unaligned in two cases: Either when no possessive expression has been used in the same syntactic construction (the *no-poss* label), or the syntactic construction of the translated sentence has been

reformulated, making the word alignment impossible (the *reword* label). Rewordings include cases where e.g. a biclausal construction in one language becomes a monoclausal construction in the other (see Example 9 in Section 5 below), comitative in one language and coordination in the other (*a boy with his father* vs. *a boy and his father*) and so on.

Table 2 shows that a significant<sup>7</sup> part of English possessive pronouns have other (or no) means of expression in Czech and Russian: Only 72% (92+80) and 82% (112+83) of English possessive pronouns are expressed by Czech and Russian possessive and reflexive-possessive pronouns, respectively. However, there is also a significant difference in expressive means between Czech and Russian. The reasons for such a difference (about 10%) can be found in the translator's style or special language features and will be addressed in more detail in later sections. It is also possible that the difference between the frequency of pronouns Russian and Czech is occasional, the lack of pronouns in Czech being partly compensated by rewording (15 in Czech vs. 10 in Russian) and externally expressed possessivity.

**Table 2.** Counterparts of English possessive pronouns in Czech and Russian

|    | poss | refl-poss | external means |       | not aligned |        |
|----|------|-----------|----------------|-------|-------------|--------|
|    |      |           | NP             | other | no-poss     | reword |
| CZ | 92   | 80        | 6              | 6     | 39          | 15     |
| RU | 112  | 83        | 5              | 3     | 25          | 10     |

**Table 3.** Counterparts of Czech possessive and reflexive possessive pronouns in English and Russian

|    | poss | refl-poss | external means |       | not aligned |        |
|----|------|-----------|----------------|-------|-------------|--------|
|    |      |           | NP             | other | no-poss     | reword |
| EN | 174  | -         | 5              | 6     | 3           | 9      |
| RU | 94   | 62        | 9              | —     | 22          | 10     |

**Table 4.** Counterparts of Russian possessive and reflexive possessive pronouns in English and Czech

|    | poss | refl-poss | external means |       | not aligned |        |
|----|------|-----------|----------------|-------|-------------|--------|
|    |      |           | NPs            | other | no-poss     | reword |
| EN | 196  | —         | 2              | 2     | 3           | 2      |
| CZ | 81   | 73        | 6              | 3     | 28          | 14     |

<sup>7</sup> This is significant at p-level  $p \leq 0.05$ . Significance has been calculated by bootstrap resampling using 100,000 samples. The same holds for all other claims about differences referred to as significant.

## 5. Analysis: English ↔ Czech & Russian: the definiteness hypothesis

One of the most interesting points addressed in Novák—Nedoluzhko (2015) concerns the expression of possessivity in English and Czech. The statistics of the correspondence of English possessive pronouns to their Czech counterparts confirms the general tendency of Czech to express personal possessive pronouns less frequently than in English. For example, in Czech, it is not common to use a possessive (or a reflexive possessive) pronoun in sentences like (4). However, it is not ungrammatical. The Czech sentence in Example 4 would remain grammatically correct after adding a reflexive possessive *svůj*.

- (4) EN: *As a result of <their> illness, they lost \$1.8 million in wages and earnings.*—  
CZ: *Důsledkem (<své>) nemoci přišli na mzdách a výdělcích o 1.8 milionu dolarů.*

We suggest that the high frequency of possessives in English is related to the grammatical category of definiteness. English has a strong tendency to avoid using bare nouns, i.e. nominal groups (especially in singular) should be mostly specified by either an article or another determiner. Possessive pronouns in cases such as *their* in Example 1 express definiteness even more explicitly than the definite article does, giving a monosemantic reference to the possessor. As a Slavic language without grammatical category of definiteness, Czech does not have such a strong tendency to express it. If our suggestion is correct, the situation in Russian should be more similar to Czech than to English, as it is also a language without obligatory expression of definiteness.

The translation of the sentence (4) into Russian supports this assumption: The use of the reflexive possessive pronoun *svoj* is grammatically correct but it is neither obligatory nor especially common.

- (4)' RU: *В результате (<своей>) болезни они потеряли \$ 1,8 млн заработной платы и других доходов.*

As shown in Table 2 for PCEDT-R, ca. 23% (39+15=54 occurrences) of English possessive pronouns are not expressed in Czech. For Russian, this number is lower: unexpressed pronouns make up ca. 15% (25+10=35 occurrences), but the difference with English is still statistically significant. In 13 cases (5%) in PCEDT-R, English possessive pronouns were not translated either into Czech or Russian. These are mostly the cases where the pronouns rather express definiteness than possessivity, cf. English *its* in *its first quarter* in Example 5.

- (5) EN: *However, Bear Stearns on Monday reported improved earnings for <its> first quarter, ended Sept. 29.*—CZ: *Společnost Bear Stearns nicméně v pondělí oznámila zvýšené výdělků za <> první čtvrtletí, končící 29. září.*—RU: *Однако, Bear Stearns в понедельник сообщил об увеличившейся прибыли за <> первый квартал, закончившийся 29 сентября.*

There are occasional examples (one found in PCEDT-R, and some were found in Intercorp or can be constructed), where an English construction with a possessive was translated into Czech with a determiner, but with a different one than used in English. See Example 6 with the demonstrative pronoun *ten* in Czech. In Russian, the reflexive possessive *svoj* remains expressed (like in English).

- (6) EN: *Lionel also urged holders of its stock and debt not to tender <their> securities...*—CZ: *Společnost Lionel též tlačí vlastníky svých akcií a dlužníky, aby <tyto> cenné papíry nenabízeli...*—RU: *Лайонел также убедил держателей его акций и долгов не номинировать <свои> ценные бумаги...*

Occasionally (4 instances in PCEDT-R), English possessives can also be translated into Czech or Russian with relative clauses (see Example 7 for Russian):

- (7) EN: *Coupled with <his> current 1.2 million shares [...] the stake would have given him control of 55% of the concern.*—CZ: *Ve spojení s <jeho> současným 1.2 milionu akcií [...] by mu tento podíl poskytl kontrolu nad 55% podniku.*—RU: *Будучи соединена с 1,2 миллионами акций, которыми он в данный момент владеет [...] эта ставка обеспечила бы ему контроль 55% концерна.*

Observing the PCEDT-R data from the perspective of Czech and Russian, we can see that much fewer possessive expressions do not find their counterparts in English, than it was for the perspective English → Czech and Russian. For Czech (see Table 3), among 197 possessive and reflexive possessive pronouns, only 12 remain unexpressed (3 *no-poss* and 9 rewordings). Other cases are expressed either with possessive pronouns (174 cases), possessive nominal groups (e.g. *the company's* instead of *its*—5 cases) or the definite article (6 cases, see Example 8). For the perspective Russian → English & Czech (see Table 4), these numbers are even smaller: Among 205 Russian possessive and reflexive possessive pronouns, 3 possessive pronouns (5%) are not expressed at all and 2 cases are rewordings. Other cases are expressed with possessive pronouns (196 cases), possessive nominal groups (2 cases) and the definite article (2 cases).

- (8) CZ: *Tento maloobchodník nebyl schopen najít pro <svoji> budovu kupce.*—EN: *The retailer was unable to find a buyer for <the> building.*—RU: *Компания была неспособна найти покупателя для <> здания.*

The interchangeability of a possessive pronoun and the definite article is especially interesting. It appears to be relatively systemic. Not only in PCEDT-R, but also in Intercorp, we can easily find examples where Czech and Russian possessive (or possessive reflexive) pronouns are aligned with the English definite article *the*. On the one hand it speaks in favor of our idea of correlation between the English grammatical category of definiteness and the frequency of possessive pronouns as compared to the Slavic languages Czech and Russian. On the other hand, it also means that

possessive pronouns in Czech and Russian do not always express the possessivity exclusively. This meaning of reflexive possessive *svoj* and other possessive pronouns was described for Russian in Brykina (2009) but as far as we know no extensive research has been done for Czech.

The analysed examples allow us to confirm the proposed definiteness hypothesis, but more data need to be analysed to formulate the distributional rules more precisely. There are also some aspects of comparison that should be addressed in more detail in further analysis. For example, it seems that in cases where possessive pronouns fill actant positions in valency frames, they seem to be more frequent in Czech and Russian than in English (Example 9). Possessives in Czech and Russian tend to be more obligatory if the possessor's antecedent is more distant from the pronouns or belongs to a different clause (*судья Ворк никогда не будет иметь высокого шанса на ?<свое>/??<его>/<>утверждение // судья Ворк никогда не будет иметь высокого шанса на то, что <его>/??<0> утверждение будет одобрено*).

Some syntactic constructions (e.g. as the strike enters <its> 15th day today at the beginning of the sentence) are typically translated from English into Czech and Russian without possessives.

- (9) CZ: *...dolu na měď ve městě Cananea v Mexiku, který byl mimo provoz od konce srpna, kdy mexická vláda oznámila <jeho> krach...*—EN: *... at the Cananea copper mine in Mexico, which hasn't been operating since <it> was declared bankrupt by the Mexican government*—RU: *...на медном руднике Кананеа в Мексике, который не действует с тех пор, как мексиканское правительство объявило о <его> банкротстве.*

## 6. Analysis: Czech ↔ Russian

### 6.1. Optionality of possessive and reflexive possessive pronouns in Czech and Russian

As we have observed in Section 4, Table 2 shows a significant difference in the frequency of possessive and reflexive possessive pronouns between Russian and Czech when translated from English possessive pronouns. It shows that possessives in Russian are used more frequently in translations. Why is it so? To answer this question, we have annotated optionality in PCEDT-R. Next to each sentence with a possessive or reflexive possessive pronoun, the label <OPT> was inserted in the cases where the possessive element could be omitted or inserted (if missing). The possibility to omit the pronoun does not necessarily infer that the meaning remains absolutely the same, it is rather our assumption that in the same pragmatic context, the sentence could be also used without this pronoun, and the possessive meaning may be mostly reconstructed from the context.

**Table 5.** Optionality of possessive means in Czech and Russian

|              | expressed OPT | unexpressed OPT | total OPT | all nodes |
|--------------|---------------|-----------------|-----------|-----------|
| cz refl_poss | 27            | 13              | 50        | 238       |
| cz poss      | 10            |                 |           |           |
| ru refl_poss | 36            | 2               | 54        |           |
| ru poss      | 16            |                 |           |           |

Table 5 shows a similar optionality in Czech and Russian:<sup>8</sup> Out of the translations of 238 English possessive pronouns, 50 and 54 cases, respectively (ca. 20% in both cases), are optional. However, in Czech, optionality was marked in a larger number of cases where possessivity was not expressed (13 cases in Czech vs. only two cases in Russian). Moreover, we observe a substantial difference in optionality of expressing possessivity between possessive and reflexive possessive in both languages: Reflexive possessives can be omitted more frequently (27 and 36 cases of reflexive possessives vs. 10 and 16 possessives in Czech and Russian respectively).<sup>9</sup> This fact appears especially interesting if we compare these numbers with the numbers in Table 2, showing that, in general, possessive pronouns are more frequently translated with possessive pronouns than with reflexive possessives in Czech and Russian.

A possessive pronoun is obligatory in cases where it fills a valency position of the inserted predication, if it expresses the contrastive meaning (e.g. *his magazines*—*many women's magazines* in Example 10) and so on.

- (10) EN: *Today, Mr. Lang believes <his> magazines will offer what many women's magazines don't.*—CZ: *Dnes je Lang přesvědčen, že <jeho> časopisy nabízejí něco, co mnohé jiné ženské časopisy nemají.*—RU: *Сегодня г-н Лэнг считает, что <его> журналы предложат то, что не предлагают многие другие женские журналы.*

The possessive meaning is often lost when omitting the possessive expressive unit. In some cases it remains clear from the context, in other cases it does not, but still we suspect that there is no special need to express it. This makes our judgement about optionality rather weak and subjective. Nevertheless, we consider it to be very important as it helps us understand the graduality of this category in Slavic languages.

Analysing examples with a different degree of optionality of possessive elements in Czech and Russian, we made some assumptions that can be confirmed or refused with a larger set of data and a more thorough analysis. For example, pronouns seem to be less obligatory when defining an inalienable part of the possessum. Also,

<sup>8</sup> This fact was not checked for statistical significance, because optionality is a subjective feature.

<sup>9</sup> The reasons for this difference can be rather empirical: reflexive possessive is coreferential to the subject, so possessor may be easier reconstructable from the context than in cases of non-reflexive possessivity. However, this has not been proved and will be eventually addressed in the future work.

in comitative constructions (A with B) a pronoun can be dropped out easier than in coordinative ones (A and B) as in Example 11.

- (11) EN: *The play concerns Teddy 's homecoming with <his> wife of six years, Ruth.*—  
CS: *Hra se soustředí na Teddyho návrat domů s manželkou Ruth, se kterou je již šest let.*—RU: *Пьеса повествует о возвращении домой Тэдди и <его> жены Рут, на которой он женился за 6 лет до того.*

## 6.2. External possession in Czech and Russian

External possession (Haspelmath, 1999) is a phenomenon where a nominal unit is syntactically encoded as a verbal dependent but semantically understood as the possessor of one of its co-arguments. Krivan (2007) claims for Czech that the variability of semantic and syntactic properties of external possession constructions is higher than in other languages of the European linguistic area. We suppose that it is also higher in Czech than in Russian. Moreover, in Czech, external possession is expressed more frequently and specific steps towards grammaticalization of this phenomenon can be observed. In Czech, possessivity is often expressed by the Dative possessor *si*, which occurred in our examples parallel to English possessive pronouns, cf. Example 12.

- (12) CZ: *Sběratelé, kteří <si> vydělali peníze na Wall Street, se stali více a více důležitou součástí obchodu s uměním*—EN: *Collectors who have made <their> money on Wall Street have become an increasingly important part of the art business.*—  
RU: *Коллекционеры, которые заработали <> деньги на Уолл-стрит, становятся всё более важной частью бизнеса искусства.*

Moreover, in Czech, there is a gradual shift between the Dative possessor *si* and the so-called free Dative that does not contain the possessive meaning anymore (see Example 13 of colloquial Czech, where three datives are used at the same time). Free Dative is quite frequent in (especially colloquial) Czech, but the borderline between possessive and non possessive meaning is not clear in many cases.

- (13) CS: *Pustila jsem dceru na hory a ona <ti> <si> <mi> tam zlomila nohu!* (Jandová 1993:62, Cit. from Krivan 2007)—[lit. EN: *I let the daughter go to the mountains, and she <to you> <her> <to me> broke the leg there].*

In Russian, cases where external possession is expressed with a reflexive pronoun are marginal in our data and it may be used in cases where it is supported by the valency frame, see Example 14:

- (14) RU: *Жители прокладывали <себе> путь через посыпанные стеклом улицы.*—EN: *Residents picked <their> way through glass-strewn streets.*—  
CS: *Obyvatelé města <si> razili cestu ulicemi zasypanými sklem.*

It appears to be reasonable to address this topic in more detail. For example, it makes sense to compare the frequency of Czech *si* with the possessive meaning in translated and original texts. We assume that in original texts the frequency will be higher, because the meaning of this particle is synthesized with more difficulties when translating from English. Making such analysis, one should also take into account the statistical significance of the output results, as well as to consider the reasonable extent of subjectivity by judging the meaning of the reflexive *si* in Czech. As was already stated in a number of theoretical studies (see e.g. Jandová 1993, Krivan 2007, etc.) the borderline between possessive and other meanings of *si* is very unclear. This analysis goes beyond the scope of the present work. Much work is yet to be done on the data, and the results exceed the limits of this paper.

As for Russian, the question of external possessivity has been analysed e.g. in Haspelmath (1999), Kibrik (2003) and Brykina (2005), but from a different perspective. Addressing external possessivity in Russian on the corpus data in comparison with Czech and English is part of our plans for the future.

### 6.3. Other differences between Czech and Russian

As we have observed in Table 2, possessive and especially reflexive possessive pronouns are significantly more frequent in Russian than in Czech texts in PCEDT-R. Heretofore, we did not find enough evidence for this difference. In order to find an explanation, we prepared a questionnaire based on the classification of meanings of the Russian *svoy* presented in Padučeva (1983) and further developed in Brykina (2009) and distributed it among Czech native speakers. We further applied our findings to the study of the parallel data in PCEDT-R. We believe that this aspect deserves a thorough analysis, but the limitations of this paper do not allow us to address all our findings in detail. In what follows, we only point out certain general areas. We found three notable groups of differences:

1. Unlike Russian, the Czech *svůj* does not have a distributive meaning. In Russian, the distributive meaning of *svoy* is frequent and is often combined with the basic possessive meaning. In such cases, Czech expresses possessivity externally, basically with the adjectives “*vlastní*” (“own”), “*jiný*” (“different”), etc.:

(15) RU: *Каждая книга оставляет в памяти свой след.*—CS: *Každá kniha zanechává v paměti jinou vzpomínku.*—[lit. EN: *Every book leaves a different trace in the memory.*] (constructed)

2. Unlike Russian, the reflexive possessive pronoun *svůj* is not common in the nominative case in Czech. In such cases, the adjective “*vlastní*” (“own”) is used:

(16) CZ: *Vlastní byt je lepší než nájemní*—RU: *Своя квартира лучше, чем съёмная.*—[lit. EN: *Own flat is better than a rented one.*] (constructed)

(17) CZ: *Vlastní zájmy mají přednost*—RU: *Свои интересы предпочтительнее.*—  
[lit. EN: *Own interests take priority.*] (constructed)

3. The use of possessive and reflexive pronouns differ in their “secondary functions”, such as the meaning close to the definite or indefinite (Example 18) article. Brykina (2009:134) also mentions a number of cases where possessive pronouns rather have a discursive than a grammatical function. Some meanings of *svoj* are close to phraseological, lose their primary possessive meaning and are expressed by other means in Czech.

(18) CZ: *Tato věta má jistou logiku.*—RU: *В этом предложении есть своя логика.*—  
[lit. EN: *This sentence follows a specific logic.*]

4. The use of reflexive possessive pronouns both in Czech and Russian has certain restrictions on the animacy of its antecedent (see e.g. Padučeva 1983 for Russian; for Czech, we did not find it described). However, these restrictions are very subtle, often allowing both variants with different stylistic or grammatical preferences.

(19) CZ: *Cestovní mapa se <svým> příslibem rychlých úspěchů [...] motivuje Palestinskou samosprávu, aby zajistila trvalou bezpečnost pro Izraelce.*—  
RU: *Путь урегулирования, с <его> обещанием быстрых результатов [...] обеспечивает стимулы для властей Палестины обеспечить израильтянам безопасность на долговременной основе.*—EN: *The roadmap, with its promise of fast achievements [...] provides incentives for the Palestinian Authority to deliver security for Israelis on a sustained basis.*

5. There is some competition between personal and reflexive possessives in syntactic constructions with embedded (explicit or implicit) predicates. As addressed in theoretical work concerning the topic (see e.g. Padučeva 1983, Panevová 1986), such cases create the potential coreferential ambiguity of the reflexive possessive pronouns, so there is a tendency to use personal possessives or other means instead. However, these tendencies may be different or they may have a different degree of obligatoriness in Czech and Russian. We have also an (yet unproven) notion that Czech has a tendency to use the reflexive possessive in embedded predicate constructions more frequently than Russian. However, this is a topic to be addressed in more detail in the future.

## 7. Translated and original texts

The fact that Czech and Russian texts in PCEDT-R have been translated from English is a relatively strong factor that can influence the frequency and distribution of possessives in the target languages. As we observed in Section 6.1, in many cases possessive pronouns are optional in Czech and Russian. However, when translating

an English sentence with explicitly expressed possessive pronouns, the translator tends to express them too, although they could potentially be omitted. Original texts are much more representative in this respect. Having the comparable Czech Prague Dependency Treebank (PDT 3.0; Bejček et al., 2013) and the Russian Treebank (RTB, Boguslavsky et al., 2000) at our disposal, we computed the frequency of possessive and reflexive possessive pronouns on a subcorpus of 1078 sentences (the same number of sentences as in PCEDT-R) in these two corpora. The results are presented in Table 6.

**Table 6:** Possessive and reflexive possessive pronouns in 1,078 sentences in original and translated texts

|                              | poss       | refl_poss |
|------------------------------|------------|-----------|
| English original (PCEDT-R)   | 238 (poss) |           |
| Czech translated (PCEDT-R)   | 92         | 80        |
| Czech original (PDT)         | 65         | 45        |
| Russian translated (PCEDT-R) | 112        | 83        |
| Russian original (RTB)       | 104        | 93        |

The results are surprising: The difference for the frequency of pronouns in original and translated texts is higher in Czech than in Russian. The situation in Russian is similar for original and translated texts, the reflexive possessive pronoun *svoj* is even slightly more frequent in original texts (93 and 83 cases respectively). On the contrary, in Czech, both possessive pronouns and *svůj* are much more frequent in translations. For possessive pronouns, there is almost a 30% difference (92 vs. 65 cases) and for *svůj* it is even higher (80 vs. 45). The most probable reason for such difference is that the Czech *svůj* lacks the distributive and some other secondary functions which are quite frequent in Russian (see Section 6.3). However, in any case, it gives us important information of another kind: Possessive elements are actually more frequent in Russian than in Czech, which is reflected in statistics for both original and translated texts.

The numbers in Table 6 also explain the observation from Section 6.1 that the Czech translation of PCEDT-R contains more cases where possessivity was not expressed but could be expressed, i.e. it was marked as optional when annotating optionality: If both options were possible – either to express possessivity or to omit it – Czech is more “economical” than Russian and tends rather not to express (especially reflexive) possessivity.

The factor of translations of the texts in our data deserves a more thorough analysis. It appears from the examples that in some cases a translated sentence contains possessive or reflexive possessive elements that would not be used in original texts, but on the other hand, there is no optionality in the narrow sense. Also, syntactic structure inducing translations with possessive pronouns exceed the limits of this paper and will be addressed in more detail in the future.

## 8. Conclusion

We have presented a corpus-based analysis of the use of possessive and reflexive possessive pronouns in the Prague Czech-English-Russian (PCEDT-R) parallel corpus. We have calculated the statistics of correspondences and analysed some tendencies that these statistics exhibit.

The created parallel data let us address differences in the expression of possessivity in the analysed languages more precisely. The statistics of pronoun correspondences in English, Czech and Russian and the interchangeability of English possessive pronouns with the definite article *the* proved the hypothesis of an existing correlation between the category of definiteness and the use of possessive pronouns. Furthermore, we analysed the differences between the use of possessive and reflexive possessive pronouns in Czech and Russian.

We believe that our findings may be interesting both from the theoretical and computational perspectives. From the perspective of computational linguistics, searching for rules of expressing possessivity helps us find and verify specific features in text that can be further used as background knowledge for the development of a multilingual tool for coreference and anaphora resolution; also, for machine translation, it is important to know which possessive pronouns should or should not be translated into Czech and Russian. From the theoretical point of view, our research contributes to contrastive comparative analysis of typologically related (Czech and Russian) and more distant (English vs. Czech & Russian) languages. The knowledge acquired by such comparison not only gives us the typologically relevant information in general but also an opportunity to know more about each separate language. For example, by comparing the specificity of use of possessive pronouns in Czech with Russian, we can understand more about each of these languages.

## 9. Acknowledgement

We acknowledge the support from the Grant Agency of the Czech Republic (grant 16-05394S). The work on this project was (partially) supported by the grant “Multilingual Corpus Annotation as a Support for Language Technologies” (LH14011, Ministry of Education, Youth and Sports). This work uses language resources developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071).

## References

1. *Bejček E., Hajičová E., Hajič J., Jínová P., Kettnerová V., Kolářová V., Mikulová M., Mírovský J., Nedoluzhko A., Panevová J., Poláková L., Ševčíková M., Štěpánek J., Zikánová Š.* (2013), Prague Dependency Treebank 3.0. Data/software, Charles university in Prague, MFF, ÚFAL, Prague, Czech Republic, <http://ufal.mff.cuni.cz/pdt3.0/>

2. *Boguslavsky I. M., Grigorieva S. A., Grigoriev N. V., Kreidlin L. G., Frid N. E.* (2000). Dependency Treebank for Russian: Concepts, Tools, Types of Information. Proceedings of the 18th Conference on Computational Linguistics. Vol 2, Saarbrücken, pp. 987–991.
3. *Brykina, M.* (2005), Types of constructions with an external possessor in Russian. [Tipy konstrukcij s vneshnim posessorom v ruskom jazyke.] Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2005” [Komp’yuternaya Lingvistika i Intellektual’nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii “Dialog 2005”].
4. *Brykina, M.* (2009), Coding the possessivity (corpus-based study of Russian) [Jazykovye sposoby kodirovanija possessivnosti (na materiale korpusnogo issledovanija ruskogo jazyka)]. Ph.D. thesis, Moscow.
5. *Čmejrková S.* (1998), Syntactic and Discourse Aspects of Reflexivization in Czech. In: Hajičová, E. (ed.): Issues of Valency and Meaning, Prague, pp. 75–37.
6. *Daneš F., Hausenblas K.* (1962), Personal and reflexive possessive pronouns in Czech [Přivlastňovací zájmena osobní a zvrtná ve spisovné češtině], *Slavica Pragensia* 4, 191–202.
7. *Dočekal M.* (2000), Reflexive possessives in Czech linguistics [Posesivní reflexivum v bohemistice], *Studia minora facultatis philosophicae universitatis brunensis*, Brno, pp. 47–59.
8. *Fed’ko E.* (2007), Non-canonical control of the reflexive pronoun in Russian [Nekanonicheskiy kontrol’ reflektivnogo mestoimenija v ruskom jazyke], Master Thesis, MSU, Moscow.
9. *Gebauer J.* (1890), Czech grammar for schools and teaching institutes [Mluvnice česká pro školy střední a ústavy učitelství], Prague.
10. *Hajič J., Hajičová E., Panevová J., Sgall P., Bojar O., Cinková S., Fučíková E., Mikulová M., Pajas P., Popelka J., Semecký J., Šindlerová J., Štěpánek J., Toman J., Urešová Z., Žabokrtský Z.* (2012), Announcing Prague Czech-English Dependency Treebank 2.0. Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), European Language Resources Association, Istanbul, pp. 3153–3160.
11. *Haspelmath M.* (1999), External Possession in a European Areal Perspective. In: Payne, D. L., Barshi, I. (eds.) *External Possession*, Amsterdam: Benjamins, pp. 109–135.
12. *Kibrik A. E.* (2003) External possession in Russian [Vneshnij possessor v ruskom jazyke]. In: Kibrik A. E. *Constants and variables in Russian* [Konstanty i peremennyye jazyka], Sankt Peterburg, pp. 307–319.
13. *Křivan J.* (2007), External possession in Czech in typological and areal perspective [Externí posesivita v češtině: v typologické a areální perspektivě], Master’s thesis. Prague, Charles University in Prague.
14. *Marcus M., Santorini B., Marcinkiewicz M.-A., Taylor A.* (1999), *Treebank-3LDC99T42*. Philadelphia: Linguistic Data Consortium. Available online: <https://catalog.ldc.upenn.edu/LDC99T42>.
15. *Novák M., Nedoluzhko A.* (2015), Correspondences between Czech and English Coreferential Expressions, *Discours: Revue de linguistique, psycholinguistique et informatique.*, Vol. 16, Caen, France, pp. 1–41.

16. *Padučeva E.* (1983), Reflexive pronoun with indirect antecedent and the semantics of reflexivity [Vozvratnoje mestoimenije s kosvennym antecedentom i semantika reflektivnosti], *Semiotika i informatika*, nr. 21, pp. 3–33.
17. *Padučeva E.* (1985). The statement and its relation to reality (referential mechanisms of the semantics of pronouns) [Vyskazyvanije i jeho sootnesennost s dejstvitelnostju (referencialnyje aspekty semantiki mestoimenij)], Moscow.
18. *Panevová J.* (1986), About reflexive pronominalisation in Russian [K voprosu o reflektivnoj pronominalizacii v češskom jazyke], *Linguistische Arbeitsberichte* 54/56, pp. 44–56.
19. *Svoboda K.* (1880), Debate on the use of personal, possessive and reflexive pronouns in shortened compound sentences [Rozprava o užívání osobných, přisvojovacích a vzratných náměstek v souvětích zkrácených], *Časopis českého muzea* 54, pp. 124–142.
20. *Testelec J., Toldova S.* (1998), Reflexive pronouns in Dagestan languages and the typology of the reflexive [Reflexivnyje mestoimenija v dagestanskix jazykax i tipologija reflexiva], *Тестелец Я. Г., Толдова С. Ю. (1998) Рефлективные местоимения в дагестанских языках и типология рефлексива. // Questions of linguistics [Voprosy jazykoznanija]*, No4, pp. 35–57.
21. *Trávníček R.* (1951), *Grammar of Czech* [Mluvnice spisovné češtiny], Prague.

## Data sources

1. *Czech National Corpus—Intercorp* [Český národní korpus—InterCorp]. Ústav Českého národního korpusu FF UK, Prague. Cit.02.02.2016, accessible from WWW: <<http://www.korpus.cz>>.
2. *Prague Czech-English Dependency Treebank* (PCEDT)—Hajič J., Hajičová E., Panevová J., Sgall P., Bojar O., Cinková S., Fučíková E., Mikulová M., Pajas P., Popelka J., Semecký J., Šindlerová J., Štěpánek J., Toman J., Urešová Z., Žabokrtský Z. (2012), Announcing Prague Czech-English Dependency Treebank 2.0. Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), European Language Resources Association, Istanbul, pp. 3153–3160.
3. *Prague Dependency Treebank* (PDT)—Bejček E., Hajičová E., Hajič J., Jínová P., Kettnerová V., Kolářová V., Mikulová M., Mírovský J., Nedoluzhko A., Panevová J., Poláková L., Ševčíková M., Štěpánek J., Zikánová Š. (2013), Prague Dependency Treebank 3.0. Data/software, Charles university in Prague, MFF, ÚFAL, Prague, Czech Republic, <http://ufal.mff.cuni.cz/pdt3.0/>
4. *Prague Czech-English-Russian Dependency Treebank* (PCEDT-R)—Novák M., Nedoluzhko A., Schwarz (Khoroshkina) A. (2016): *planned to be published in Lindat/Clarin before the camera-ready version of this paper.*
5. *Russian Treebank* (RTB). Accessible from WWW: [otipl.philol.msu.ru/~soiza/rtb/](http://otipl.philol.msu.ru/~soiza/rtb/)