Grammatical number of nouns in Czech: linguistic theory and treebank annotation

Magda Ševčíková, Jarmila Panevová, Zdeněk Žabokrtský

Charles University in Prague, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
E-mail: {sevcikova,panevova,zabokrtsky}@ufal.mff.cuni.cz

Abstract

The paper deals with the grammatical category of number in Czech. The basic semantic opposition of singularity and plurality is proposed to be enriched with a (recently introduced) distinction between a simple quantitative meaning and a pair/group meaning. After presenting the current representation of the category of number in the multi-layered annotation scenario of the Prague Dependency Treebank 2.0, the introduction of the new distinction in the annotation is discussed. Finally, we study an empirical distribution of preferences of Czech nouns for plural forms in a larger corpus.

1 Introduction

Morphological categories are described from formal as well as semantic aspects in grammar books of Czech (and of other languages; for Czech see e.g. [7]). Both these aspects are reflected in the annotation of Prague Dependency Treebank version 2.0 (PDT 2.0; see [4] and `http://ufal.mff.cuni.cz/pdt2.0`). In the present paper, the grammatical category of number of nouns is put under scrutiny.

In Section 2, we explain the basic semantic opposition of the category of number (singularity vs. plurality, prototypically expressed by singular and plural forms, respectively) and focus on nouns that are used predominantly, or even exclusively, either in singular or in plural. Our analysis is based on large corpus data and confronted with the traditional linguistic terms (such as singularia tantum, pluralia tantum). We propose to enrich the basic singular-plural opposition with the opposition of simple quantitative meaning (concerning number of entities) vs. pair/group meaning (number of pairs/groups).

Current representation of formal and semantic features of the category of number within the multi-layered annotation scenario of PDT 2.0 is briefly introduced in Section 3. The PDT 2.0 annotation scenario has been designed on the theoretical basis of Praguian Functional Generative Description (FGD; [14]). As the linguistic

theory of FGD has been still elaborated and refined in particular aspects, the question of the introduction of the recent theoretical results is to be asked. Section 4 brings novel observations related to the distribution of preferences of individual nouns for plural forms derived from the Czech National Corpus data.

2   The category of number in Czech

2.1   Meaning and form of the category of number

Number is a grammatical category of nouns and other parts of speech in Czech. With nouns, the category of number reflects whether the noun refers to a single entity (singularity meaning) or to more than one entity (plurality). With other parts of speech (adjectives, adjectival numerals, verbs), the number is imposed by agreement. In this paper we deal with the number of nouns only.[1]

In Czech, the number is expressed by noun endings. Czech nouns have mostly two sets of forms directly reflecting the opposition of singularity and plurality: singular forms and plural forms. Nouns *ruka* 'arm', *noha* 'leg', *oko* 'eye', *ucho* 'ear', *rameno* 'shoulder', *koleno* 'knee' and diminutives of these nouns have an incomplete set of (historical residues of) dual forms as well, which are used instead of the plural forms when referring to body parts.

According to the data from the SYN2005 subcorpus of the Czech National Corpus,[2] singular and plural forms of nouns occur roughly in the ratio 3:1 in Czech texts (22,705,247 singular forms:7,440,382 plural forms). Concerning the ratio of singular and plural forms for single nouns, a detailed analysis of the SYN2005 data was carried out. The SYN2005 corpus contains 452,015 distinct noun lemmas, only those with more than 20 occurrences were involved in the analysis (48,806 lemmas). The majority of Czech nouns (42,550 lemmas out of 48,806) is used in singular more often than in plural (see Fig. 1 (a); for further details see Sect. 4).[3] At both ends of the scale there are nouns that clearly prefer either singular or plural forms, or are even limited to the singular on the one hand or to the plural on the other.

In our opinion, the predominance of singular or plural forms can be traced back, roughly speaking, to two factors. The first of them is the relation of the noun to the extra-linguistic reality: some nouns refer to objects that occur in the reality mostly separately or in large amounts, respectively. The other one lies in the language itself, more specifically, in the process of structuring the described reality by the language: for instance, groups of some entities are considered as a

---

[1]According to Mathesius ([8]), the number of nouns belongs to functional onomatology, with other parts of speech it is a part of functional syntax.

[2]SYN2005 is a representative corpus of Czech written texts, containing 100 million both lemmatized and morphologically tagged tokens ([2]).

[3]The ratio between singular and plural forms corresponds to the known language principle considering the singular to be the unmarked member of the basic number opposition, it can be used in some contexts instead of its marked counterpart (plural).

single whole and as such referred to by the singular form (e.g. so-called collective nouns); on the contrary, some single objects are, often due to their compoundness or open-endess, referred to by the plural form (so-called pluralia tantum).

Nouns preferring singular or plural forms are discussed in Sect. 2.2 and 2.3, in Sect. 2.4 a new semantic distinction is introduced.

2.2 Nouns preferring singular forms

More than a third of the analyzed noun lemmas (16,473 lemmas out of 48,806) was used exclusively in singular in the SYN2005 data. Most of them are proper names (12,286 lemmas; see Fig. 1 (a) and (b)).[4] Only singular forms were found also for nouns such as *dostatek* 'sufficiency', *údiv* 'astonishment', *sluch* 'hearing', *kapitalismus* 'capitalism', *zámoří* 'overseas', *severovýchod* 'northeast', *potlačování* 'repression', *pohřbívání* 'burying', *arabština* 'Arabic'. A strong preference, though not exclusivity, of singular forms is characteristic for nouns denoting a person, an object, an institution, en event etc. that is unique or fulfils a unique function in the given context or segment of reality, e.g. *svět* 'world', *republika* 'republic', *prezident* 'president', *začátek* 'beginning', *centrum* 'center' (see e.g. uniqueness of *Česká republika* 'Czech Republic', *prezident USA* 'President of the U.S.').

In case of proper names, the predominance of singular forms can be seen as anchored in the extra-linguistic reality (see Sect. 2.1): as they refer to a person, an object, a place etc. and identify them as individuals, they often occur in singular. However, for the absolute majority of Czech proper names plural forms can be formed, they are used to refer to several persons, objects etc. named with the same proper name (see the plural of the first name *František* in ex. (1)) or metaphorically (ex. (2)) etc. The other (i.e. intralinguistic) "reason" for the preference of singular concerns collective nouns (e.g. *dělnictvo* 'labour', *hmyz* 'insects', *listí* 'leaves').

Proper names, collectives as well as mass nouns (*voda* 'water', *měd'* 'copper'), names of processes and qualities (*kvašení* 'fermentation' and *sladkost* 'sweetness') and possibly other are subsumed under the term of singularia tantum in grammar books of Czech (e.g. [7]). Since we have found out that plural forms of proper names, collectives etc. do occur in the corpus data (though with a lower, but not insignificant frequency) we consider the term singularia tantum rather inappropriate. Beside this, in grammars of Czech the scope of this class is defined vaguely. The potentiality of the grammatical system to form a full paradigm with both singular and plural forms opens the possibility to use these plural forms for meaning shifts, metaphors, occasionalisms etc. (see ex. (3) and (4)).[5]

---

[4]There were nearly 20,000 proper name lemmas with 20 or more occurrences found in the SYN20005 data. More than 12,000 of them occurred in singular only, more than 900 had only plural forms (Sect. 2.3), more than 7,000 proper name lemmas were used both in singular and plural.

[5]The other way round, when considering the nouns that are truly limited to singular in the corpus data, they have no noticeable (semantic, derivational etc.) feature in common and to cover them with the term singularia tantum seems not to be profitable.
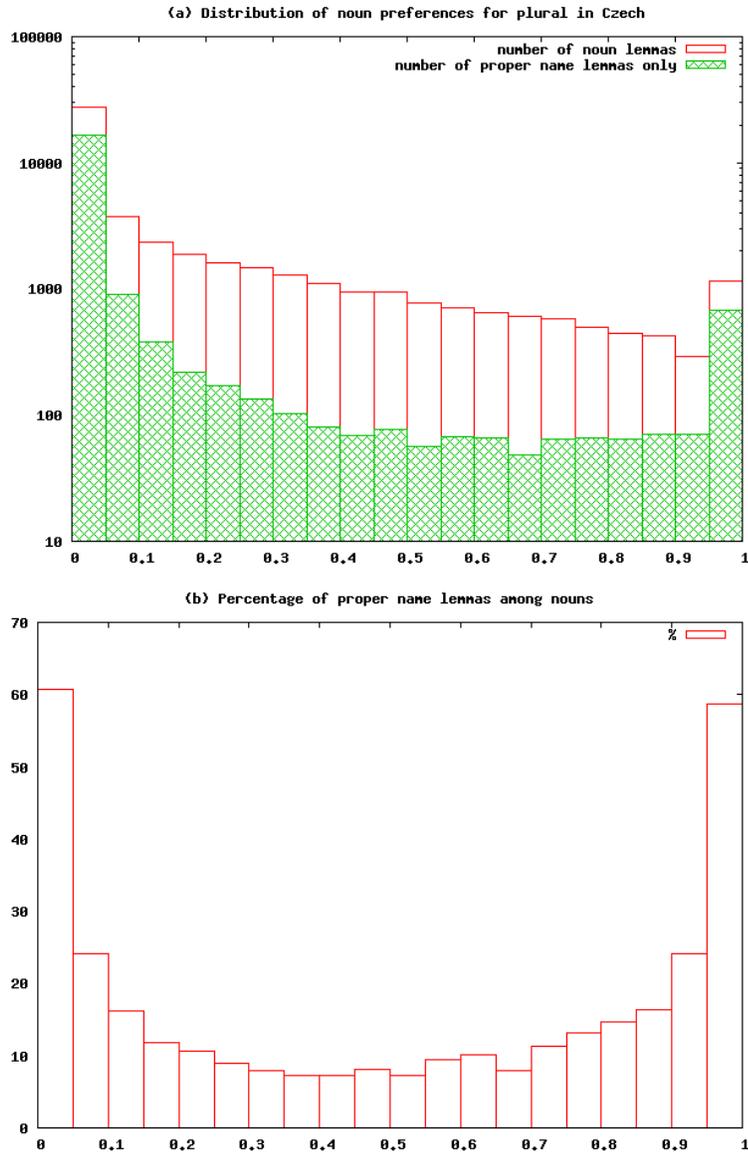
Figure 1: (a) Histogram of noun preferences for plural according to the SYN2005 corpus. The horizontal axis represents the ratio of plural forms (among all occurrences of a noun lemma), the vertical axis represents the number of distinct lemmas having its preference for plural in a given interval. (b) Percentage of proper names among noun lemmas with respect to their preferences for plural.

(1) *Třeba na Tomáše Kaberleho čekali dva **Františkové**, otec i bratr.* 'For instance, two **Františeks**, father and brother, were waiting for Tomáš Kaberle.' (SYN2005)

(2) *Vymizí **Goethové**, také **Beethovenové** se ztratí či jsou umlčeni.* '**Goethes** disappear, **Beethovens** get lost or will be silenced as well.' (SYN2005)

(3) *Čerstvé listy špenátu očistíme a propláchneme alespoň ve třech **vodách**.* 'Fresh spinach leaves are to be cleaned and washed in at least three **waters**.' (SYN2005)

(4) *Nikdy neodolá **sladkostem**.* 'He never resists **sweets**.' (SYN2005)

2.3 Nouns used predominantly in plural

The preference of plural is typical for much fewer nouns than the preference of singular; only 941 lemmas (out of 48,806 noun lemmas analyzed) occurred more often in plural than in singular in the SYN2005 corpus data. Most of these nouns are proper names (607 lemmas; in particular toponyms) and nouns such as *záda* 'back/backs', *noviny* 'newspaper/newspapers', *vrata* 'gate/gates', which are referred to as pluralia tantum in linguistic terminology. The set of forms of most pluralia tantum is truly limited to plural forms, the plural is used to refer to a single object as well as to a number of them (the current meaning is to be resolved on the basis of context, knowledge of situation etc.; cf. the noun *dveře* 'door/doors' in ex. (5)). A singular form of a plurale tantum is used only exceptionally (e.g. with *kalhoty* 'trousers', *brýle 'glasses'*), nevertheless, it has usually a shifted meaning (*kalhota* as *nohavice* 'trouser leg') and the plural preserves the ambiguity. The term of pluralia tantum proved to be adequate according to the performed data analysis (it is used in the present paper).

Besides proper names and pluralia tantum, among nouns with only plural forms there were only few nouns that have both singular and plural forms, e.g. *arašíd* 'peanut', *monocyt* 'monocyte', *autodíl* 'spare part', *johanita* 'Knight of Malta'. In our opinion, the singular of these nouns, though commonly available, did not occur in the data since the nouns refer to persons, objects etc. that usually occur in groups or large numbers in the reality. The same reason applies to nouns such as *hasič* 'fireman', *potravina* 'food', *živina* 'nutrient', *dohad* 'guess' that occurred in singular in less than 10 % of their corpus occurrences.

(5) *Z chodby byly otevřené jen **dveře** do kuchyně, ostatní **dveře** byly zavřené.* 'In the corridor, only the **door** to the kitchen was open, the other **doors** were closed.'

2.4 Pair nouns and group nouns

It is worth noting that some nouns with the plural preference do not refer just to a larger amount of entities but prototypically to a pair or to a usual group or set of them – we speak about pair/group nouns.[6] For instance, the plural form *ruce* 'hands' means usually the pair of the upper limbs, not just a larger amount of

---

[6]Vossen and Copestake [16] use the term "group noun" for nouns such as *band* that refer to a group of people etc.

them, *rodiče* 'parents' are usually a pair of persons (mother and father), not several mothers and/or fathers, *vlasy* 'hair' is a (mostly not precisely quantifiable) set of hairs on one's head, *klíče* 'keys' refer to a bunch of keys. Besides the pair/group meaning the plural form of these nouns can express, without any formal change, a larger number of pairs/groups of the objects in question or simply a larger number of the objects (common plural meaning) as well.[7]

The difference between the meanings of a pair/group and several pairs/groups on the one hand and the common plural on the other becomes evident when counting the objects: in Czech the noun that refers to a pair/group (or to several pairs/ groups) is compatible only with a set numeral, while when used in the common plural meaning, the amount is expressed by a cardinal; see ex. (6) and (7). Set numerals ("souborové číslovky" in Czech terminology) are a special subtype of numerals expressing the number of pairs and other groups.[8] Set numerals are available, for instance, in Serbian and Croatian as well whereas in English or German they have no counterpart within numerals, the number of sets is then indicated by expressions such as *a pair/two pairs*. In Czech the form of numerals is one of the means for resolving the ambiguity between the common plural meaning and the pair/group meaning.

For the already mentioned nouns *ruce*, *vlasy*, *klíče*, *rodiče*, *boty*, *rukavice* and many others, the pair/group meaning is frequent, though not limited to them. It could be, according to the recent linguistic analysis based on large corpus data [12], expressed by most Czech concrete nouns. The hypothesis that the pair/group meaning is not bound up with nouns as lexical units is supported, for instance, by the unlimited co-occurrence of nouns with set numerals (ex. (8)) or by the fact that each noun which expresses the pair/group meaning in a particular context can be used in the common plural meaning (or in singular expressing singularity) in other contexts.

Therefore, we propose to consider the pair/group meaning as a semantic feature opposed to the simple quantitative meaning. Combining this distinction (set vs. simple) with the basic singular-plural opposition, four combinations are to be considered: sg.simple (singularity meaning), sg.set (meaning of a pair/group), pl.simple (common plural meaning), and pl.set (meaning of several pairs/groups). With nouns having both singular and plural forms, the meaning sg.simple is expressed by the singular, the other three meanings by the plural form. With pluralia tantum, a plural form is used for all four meanings.

(6) *Během pár týdnů jsem protančila **troje boty***. 'During a few weeks I wore through **three pairs of**

---

[7]Dual forms (available for the above listed nouns) are not distinguished from plural forms here since they just discern the body part meaning from the other meanings of the particular noun (e.g. the instrumental dual form *očima* for human eyes vs. instrumental plural form *oky* for loops in mesh) but have no distinctive function concerning the pair/group meaning in Czech: dual forms (just as plural forms with other nouns) refer to a pair or several pairs of the particular body part as well as to a large amount of them.

[8]In connection with a plurale tantum, a set numeral expresses the number of pieces of entities.

shoes by dancing.' (SYN2005)

(7) *Až doma zjistil, že mu prodali **dvě** levé **rukavice***. 'Only after his arrival home he found out that they had sold **two** left **gloves** to him.'

(8) *Najdeme-li **dvoje** velké **stopy** a mezi nimi **jedny** menší, řekneme si: "rodina na výletě"*. 'If we find **two sets of** big **tracks** and **one set of** smaller **ones** between them, we say: "a family on a trip".' (SYN2005)

3 The category of number in the multi-layered annotation scenario of PDT 2.0

3.1 Annotation of number at the morphological layer of PDT 2.0

Formal morphological characteristics of words are described at the morphological layer of PDT 2.0. At this layer, a positional tag specifying part of speech and particular morphological categories was assigned to each token. The fourth position of the tag is reserved for the category of number. With noun forms and forms of other parts of speech that are marked for number (adjectives etc.), one of five values is to be assigned: basic values S and P with singular and plural forms, respectively, the value D with dual forms, values W and X for ambiguous cases; see [5].[9]

The preference of nouns either for singular or for plural is reflected in the morphological annotation by the assignment of plural lemmas to pluralia tantum. The set of nouns with plural lemmas roughly corresponds with lemmas marked as pluralia tantum in representative dictionaries of Czech ([6], resp. [3]); a singular form of a plurale tantum (such as *kalhota*, see Sect. 2.3) is assigned a plural lemma (*kalhoty* in the respective case) and the value S at the fourth tag position. However, neither pluralia tantum nor nouns limited to singular are marked explicitly in the morphological annotation.

3.2 Annotation of number at the tectogrammatical layer of PDT 2.0

The meaning of morphological categories is involved in the so-called tectogrammatical annotation of PDT 2.0, at which the (linguistic) meaning of the sentence is described as a dependency tree structure consisting of labeled nodes and edges.[10] At the tectogrammatical layer, the meaning of the category of number is encoded in the grammateme number. Grammatemes are node attributes capturing the meaning of semantically relevant morphological categories such as number and gender for nouns, degree of comparison for adjectives and adverbs, tense and aspect for verbs.[11] The grammateme number was assigned to nouns and substantival pro-

---

[9]With parts of speech that do not express number (e.g. prepositions), a dash (-) was filled in.

[10]Besides the morphological and tectogrammatical annotation, PDT 2.0 data were assigned also at the so-called analytical layer. At this layer, a dependency tree describing the surface-syntactic structure was assigned to each sentence.

[11]On the contrary, e.g. neither the category of case for nouns nor that of gender for adjectives were captured within the tectogrammatical annotation as they are only imposed by government or agree-

nouns and numerals (for details see [10], [13]).

Two values of the grammateme number were defined: sg and pl. Since the majority of Czech nouns express the semantic opposition of singularity and plurality directly, the values of the number grammateme could be assigned mostly automatically, mapping the number value involved in the morphological tag onto the grammateme value: the tag value S corresponds to the grammateme value sg, tag values P and D to the grammateme value pl.[12]

From the nouns described in Sect. 2.2 and 2.3, the grammateme value was assigned manually only to nouns that strongly prefer plural and use this form to refer to a single entity as well as to a larger amount of them (pluralia tantum). Since pluralia tantum were not marked explicitly in the PDT 2.0 data nor e.g. in the morphological dictionary used for tagging,[13] the manual annotation concerned nodes whose lemma was used in plural in more than 95 % lemma occurences in the PDT 2.0 data and nodes with lemmas that were marked as pluralia tantum in the Dictionary of Standard Czech Language ([6]).[14]

3.3 Annotating the pair/group meaning

The pair/group meaning explained in Sect. 2.4 has been introduced in the theoretical background only recently ([12])[15] and was not involved in PDT 2.0. Nevertheless, since the annotation scenario of PDT has been built on the theoretical basis of FGD, reflecting the state-of-the-art of this framework, and currently a new, both revised and extended, version of PDT (PDT 3.0) is being prepared (the revision concerns annotation of grammatical categories as well, see [11]), we are facing the question whether the pair/group meaning should be incorporated in the PDT 3.0 annotation scenario.

Before any large-scale annotation can start, it should be checked whether the phenomenon to be annotated (the pair/group meaning of plural) is reasonably frequent and practically distinguishable in the data. We performed the following pilot annotation experiment. Within 1,000 plural forms randomly selected from the SYN2005 corpus, the pair/group meaning was identified in 55 cases. If we project the same ratio on the tectogrammatically annotated sections of PDT 2.0 (which contain around 60,000 occurrences of denotative nouns in plural forms), we could

---

ment, respectively. Treatment of grammatical categories in FGD is closely related to the approach of Meaning–Text Theory (cf. [9]; correspondences between FGD and MTT are analyzed in [18]).

[12]Nouns, substantival pronouns and numerals with tag values W and X were assigned manually with sg or pl according to their meaning or with the value nr defined for semantically ambiguous cases. A special value inher was assigned to reflexive and relative pronouns that "inherit" the number from the coreferred node (in cases of grammatical coreference).

[13]Unlike the PDT 2.0 data and tools, information on pluralia (as well as singularia) tantum is involved, for instance, in the Croatian Morphological Lexicon [15].

[14]The lemma list obtained from PDT 2.0 data overlapped with that extracted from [6] to a large extent.

[15]In spite of the fact that this semantic feature was mentioned already in [7]. Some remarks concerning the way of expressing the pair/group meaning in Hungarian, Brazilian Portuguese, Syrian Arabic or Dutch can be found in [1].

expect roughly three thousand occurrences of the pair/group meaning. This seems to be a sufficiently high frequency: if we compare it to the frequency of functor values (i.e. dependency relations, semantic roles) annotated at the tectogrammatical layer, more than half of them does not reach this number (e.g. the functor HER for modifications with the meaning of heritage or TFRHW for modifications with the temporal meaning "from when").

Before starting the large-scale annotation, it will be further necessary to measure the inter-annotator agreement and to find ways how to automatically exclude plural forms that are not likely to have the pair/group meanings, so that the set of annotation instances is maximally reduced.

4   Empirical distribution of preference for plural

As we have already mentioned above, the average ratio of occurrences of singular and plural noun forms in Czech texts is 3:1. Obviously, nouns largely differ in their preferences for singular and plural. This section investigates the distribution of such preferences over the vocabulary of Czech nouns. For the purpose of this experiment, we ignore the fact that singular/plural preferences may vary across different senses of a single noun.

Let us have a function $pl(l)$ which expresses the preference of a noun lemma l for plural forms simply as a relative proportion of occurrences of plural word forms among all tokens with the lemma l in a given corpus (in other words, it estimates the probability of plural given the lemma).

We would like to estimate the distribution of values of $pl(l)$ across the noun vocabulary. Instead of PDT 2.0, which is too small for such estimates, we used SYN2005, which contains 100 million tokens. There are around 450,000 distinct noun lemmas in SYN2005, with around 30 million occurrences in total, out of which 7.4 million are plural forms. We divided the range of values of $pl(l)$ uniformly into 20 subintervals. We disregarded lemmas with less than 20 occurrences. Fig. 1 (a) shows the resulting histogram with the vertical axis representing the number of distinct noun lemmas having $pl(l)$ within a given subinterval.

One can immediately see two peaks in the leftmost and rightmost subintervals. The inner part of the histogram with $pl(l)$ between 0.1 to 0.9 resembles an exponential distribution. This is visually emphasized by using the logarithmic scale on the vertical axis, as the curve becomes close to linear (note that the same pattern can be seen in the distribution of English noun preferences for plural derived from the British National Corpus, see Fig. 2).

This is a striking observation. First, the distribution for values between 0.1 and 0.9 seems to be monotonous. If there were no assumptions about the $pl(l)$'s distribution across the noun vocabulary, one would expect rather a bell-shaped curve with the peak close to the average value of $pl(l)$, which is 0.16.[16] Second, the fact

---

[16]Recall that the histogram shows how the vocabulary of noun lemmas is partitioned with respect to their preference for plural, regardless of their total frequency in the corpus (the frequency was
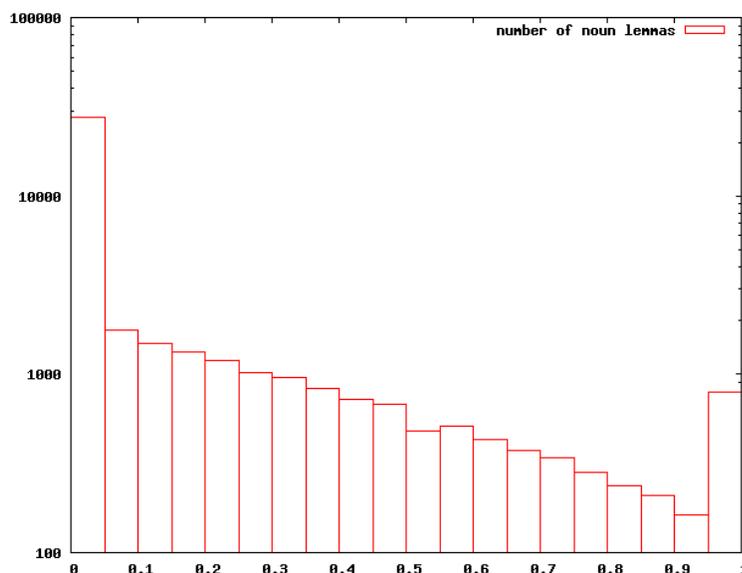
Figure 2: Histograms of noun preferences for plural according to British National Corpus.

that the distribution seems to be so regular and quite close to a perfect exponential shape suggests that there must be some relatively simple control mechanism in the language and that the distribution reflects a (dynamic) equilibrium to which this mechanism has led. The mechanism could be related to the process of lexical diversification during the language evolution.[17]

In our opinion, the key to the distribution lies in the language economy: the language tends to minimize the "energy" needed for expressing a meaning to be conveyed wherever possible. Expressing plural forms is usually more demanding than expressing singular forms. So always when a plural form of a certain noun lemma is used, the speaker might be "tempted" (not on the conscious level and not very intensively, though) to introduce a new word which could express a similar meaning by a singular form (e.g. *forest* instead of *trees*). If the new word gets spread over the population of language users, it will partially substitute the original noun and thus the original noun's $pl(l)$ will decrease. A dynamic equilibrium between this force and forces in the opposite direction (increasing the vocabulary is also costly) is reached. The fact that the distribution has an exponential form

---

used only for pruning infrequent nouns for which the estimate would be too unreliable). That is why the average $pl(l)$ does not correspond to the proportion of plural forms in the corpus.

[17]We do not expect the distribution to be predominantly influenced by extra-linguistic factors. It is difficult to imagine any language-independent prior distribution of singularity versus repetitiveness in the physical world around us; the distinction depends rather on how we structure our perception of the world by our language.

suggests that the equilibrium can be described by a first-order linear differential equation. However, this is only a preliminary hypothesis that should be further elaborated using laws of quantitative linguistics ([17]) and verified on other grammatical oppositions.

5   Conclusion

The paper is focused on the grammatical category of number of nouns within the theoretical linguistic description and the annotation of PDT 2.0. Based on large corpus data analysis, special attention has been paid to nouns with strong preferences either for singular or plural forms.

Besides the quantitative observations, the semantic opposition of singularity and plurality, which constitutes the category of number, has been refined with the distinction of the simple quantitative meaning and the pair/group meaning. The inclusion of the established opposition in the annotation scenario is not surprising and needs not to be justified whereas the involvement of the newly proposed distinction of the simple quantitative meaning and the pair/group meaning is to be carefully discussed.

Acknowledgements

References

[1] Corbett, G. G. (2000) *Number*. Cambridge University Press, Cambridge.

[2] Czech National Corpus - SYN2005. Institute of Czech National Corpus, Faculty of Philosophy and Arts, Charles University in Prague, Prague 2005. <http://www.korpus.cz>

[3] Filipec, J. et al. (1998) *Slovník spisovné češtiny pro školu a veřejnost*. Academia, Praha.

[4] Hajič, J. et al. (2006) *Prague Dependency Treebank 2.0*. CD-ROM, Cat. Nr. LDC2006T01. Linguistic Data Consortium, Philadelphia.

[5] Hana, J. et al. (2005) *Manual for Morphological Annotation. Revision for the Prague Dependency Treebank 2.0*. ÚFAL Technical Report No. 2005/27. ÚFAL MFF UK, Prague.

[6] Havránek, B. et al. (1989) *Slovník spisovného jazyka českého*. 2. vydání. Academia, Praha.

[7] Komárek, M. et al. (1986) *Mluvnice češtiny 2*. Academia, Praha.

[8] Mathesius, V. (1929) Funkční lingvistika. In *Sborník přednášek pronesených na Prvém sjezdu československých profesorů filosofie, filologie a historie v Praze 3.–7. dubna 1929*, pp. 118–130. Praha.

[9] Mel'čuk, I. A. (1988) *Dependency Syntax: Theory and Practice*. State University of New York Press, New York.

[10] Mikulová, M. et al. (2006) *Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual*. ÚFAL Technical Report No. 2006/30. ÚFAL MFF UK, Prague.

[11] Panevová, J. and Ševčíková, M. (2010) Annotation of Morphological Meanings of Verbs Revisited. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, pp. 1491–1498. ELRA, Paris.

[12] Panevová, J. and Ševčíková, M. (in prep.) Počítání substantiv v češtině (Poznámky ke kategorii čísla). To appear in *Slovo a slovesnost*.

[13] Razímová, M. and Žabokrtský, Z. (2006) Annotation of Grammatemes in the Prague Dependency Treebank 2.0. In *Proceedings of the LREC 2006 Workshop on Annotation Science*, pp. 12–19. ELRA, Paris.

[14] Sgall, P., Hajičová, E. and Panevová, J. (1986) *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. D. Reidel Publishing Company, Dordrecht.

[15] Tadić, M. and Fulgosi, S. (2003) Building the Croatian Morphological Lexicon. In *Proceedings of the EACL2003 Workshop on Morphological Processing of Slavic Languages*, pp. 41–46. ACL, Budapest.

[16] Vossen, P. and Copestake, A. (1993) Untagling Definition Structure into Knowledge Representation. In Briscoe, T., de Paira, V and Copestake, A. (eds.) *Inheritance, Defaults and the Lexicon*, pp. 246–274. Cambridge University Press, Cambridge.

[17] Wimmer, G. and Altmann, G. (2005) Unified derivation of some linguistic laws. In *Quantitative Linguistics*, pp. 791–807. De Gruyter, Berlin & New York.

[18] Žabokrtský, Z. (2005) Resemblances between Meaning–Text Theory and Functional Generative Description. In *Proceedings of the 2nd International Conference of Meaning–Text Theory*, pp. 549–557. Slavic Culture Languages Publishers House, Moscow.