



---

The Prague Bulletin of Mathematical Linguistics

NUMBER 96 OCTOBER 2011 27-48

---

## Specificity of the number of nouns in Czech and its annotation in Prague Dependency Treebank

Magda Ševčíková, Jarmila Panevová, Lenka Smejkalová

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

---

### Abstract

The paper focuses on the way how the grammatical category of number of nouns will be annotated in the forthcoming version of Prague Dependency Treebank (PDT 3.0), concentrating on the peculiarities beyond the regular opposition of singular and plural. A new semantic feature closely related to the category of number (so-called pair/group meaning) was introduced. Nouns such as *ruce* 'hands' or *klíče* 'keys' refer with their plural forms to a pair or to a typical group even more often than to a larger amount of single entities. Since pairs or groups can be referred to with most Czech concrete nouns, the pair/group meaning is considered as a grammaticalized meaning of nouns in Czech. In the present paper, manual annotation of the pair/group meaning is described, which was carried out on the data of Prague Dependency Treebank. A comparison with a sample annotation of data from Prague Dependency Treebank of Spoken Czech has demonstrated that the pair/group meaning is both more frequent and more easily distinguishable in the spoken than in the written data.

---

### 1. Introduction

In Czech, nouns typically have two sets of forms according to the grammatical category of number: singular forms and plural forms. Forms of the former set are used to denote a single entity (singularity meaning, *sg*), plural forms express, in general, more than one entity (plurality meaning, *pl*).

In addition to the existence of nouns accompanied in the lexicon with the feature "singulare tantum", which blocks the semantic opposition of *sg* vs. *pl*, and "plurale tantum", where the opposition of *sg* and *pl* is expressed by the same form, we introduce a new semantic feature closely related to the category of number, namely

the “pair/group meaning” (Section 2 of the paper). Nouns such as *ruce* ‘arms’, *boty* ‘shoes’ or *klíče* ‘keys’ refer with their plural forms rather to a pair or to a typical group even more often than to a larger amount of single entities; thus the plural *ruce* denotes a pair or several pairs of arms rather than several upper limbs, the form *boty* ‘shoes’ usually denotes a pair or several pairs of shoes, the form *klíče* ‘keys’ often means a bundle or more bundles of keys. Since pairs or groups can be referred to with most Czech concrete nouns and since this phenomenon is reflected in some peculiarities as to the compatibility of the particular nouns with numerals, the pair/group meaning is considered as a grammaticalized meaning of nouns in Czech.

In Section 3 of the paper, the manual annotation of the pair/group meaning is described, which was carried out on the data of Prague Dependency Treebank version 2.0 (PDT 2.0)<sup>1</sup> by two human annotators in parallel. The annotation was evaluated in several aspects (inter-annotator agreement, frequency of the pair/group meaning with particular nouns etc.).

Results of the annotation of the written data from PDT 2.0 are compared with a sample annotation of the data from Prague Dependency Treebank of Spoken Czech. The fact that the pair/group meaning is both more frequent and more easily distinguishable in the spoken than in the written data is briefly discussed in Section 4. The annotation of the pair/group meaning is to be included in the forthcoming version of Prague Dependency Treebank (PDT 3.0), which is designed as a both revised and extended version of PDT 2.0 (Sect. 5).

## 2. The pair/group meaning of Czech nouns

### 2.1. Nouns expressing pairs or groups

The starting point of the considerations on the pair/group meaning was an analysis of Czech nouns the plural of which usually refers to pairs or groups of entities, not to a plurality of single entities, though they are countable as single entities and also the regular opposition of *sg* and *pl* is applicable here (*jeden klíč* ‘one key’, *dva klíče* ‘two keys’ etc.). This phenomenon concerns especially nouns denoting body parts occurring in pairs or groups (*uši* ‘eyes’, *prsty* ‘fingers’, *vlasy* ‘hair’), further clothes and accessories for these body parts (*náušnice* ‘earrings’, *rukavice* ‘gloves’), family members such as *rodiče* ‘parents’, *sourozenci* ‘siblings’, and objects of everyday use and foods sold or used in typical amounts (*klíče* ‘keys’, *sirky* ‘matches’, *cigarety* ‘cigarettes’, *sušenky* ‘biscuits’).

Plural forms of other Czech concrete nouns may refer to pairs or groups of entities as well but, according to a detailed corpus analysis, they are mostly accompanied with a so-called set numeral in such contexts. Set numerals are considered to be a

---

<sup>1</sup>See (Hajič et al., 2006), <http://ufal.mff.cuni.cz/pdt2.0>

special sub-type of numerals in Czech (besides the cardinal, ordinal etc. numerals),<sup>2</sup> classification of numerals as set numerals is based on their formal shape, not on their meaning; set numerals are compatible with nouns in plural only.<sup>3</sup> The primary meaning of the set numerals is to express different sorts of the entities denoted by the noun (ex. (1)). However, the same set numerals, if combined with pluralia tantum nouns, express either the amount of single entities (i.e. the same meaning which is expressed by cardinal numerals with most nouns), or the number of sorts, cf. ex. (2). The set numerals in combination with the nouns which we are interested in in the present paper express the number of pairs or groups; this means that the set numerals are used here instead of cardinal numerals while the cardinals combined with these nouns express the number of single entities (cf. *dvoje boty* ‘two pairs of shoes’, *troje boty* ‘three pairs of shoes’ vs. *dvě boty* ‘two shoes’, *tři boty* ‘three shoes’).<sup>4</sup>

- (1) *Máme dvoje sklenice – na bílé a červené víno.*<sup>5</sup>  
‘We have **two sets** of glasses – for the white and for the red wine.’
- (2) *Na stole leží troje nůžky.*  
‘There are **three types/pieces** of scissors on the table.’

Due to the ambiguity of the set numerals as well as to the fact that pairs or groups are referred to mostly by nouns that are not accompanied with a set numeral, the pair/group meaning is not attributed to the numerals, it is proposed to be considered as a meaning of nouns. If a noun denoting a pair or group collocates with a numeral, the meaning of the noun is only reflected in the surface form of the numeral, i.e. the set numeral is used.

## 2.2. The pair/group meaning as a grammatical meaning

As we aimed at including the pair/group meaning into the theoretical description of the Czech language, namely into the framework of Functional Generative Descrip-

<sup>2</sup>Set numerals are available, for instance, in Russian, Serbian and Croatian as well; however, there are several differences in counting pairs and groups between these languages and Czech. In English and German, on the other hand, there are no numerals of this type, the number of pairs and groups is expressed by cardinal numerals in combination with the nouns such as *pair*, *bundle* and *Paar*, *Bündel*, respectively (cf. the English translations of the examples given in the text); see (Panevová and Ševčíková, 2011).

<sup>3</sup>Well established terms for formal and semantic aspects of numerals, suitable also for covering such irregularities, are missing in Czech linguistics.

<sup>4</sup>Within the tectogrammatical annotation of PDT 2.0 the numerals of both types *tři* and *troje* are represented by a single “deep” lexical item and the particular (cardinal, set etc.) meaning is represented as a separate semantic feature (grammateme numertype) according to the meaning of the counted noun and to the current context; see Fig. 2 in the paper, further details can be found in (Razímová and Žabokrtský, 2006).

<sup>5</sup>Examples (1) to (4), (16) and (17) were created as illustrative examples by the authors, all the remaining examples come from the PDT 2.0 data.

tion (FGD; (Sgall, 1967), (Sgall et al., 1986)) and possibly also into the annotation of PDT, the annotation scenario of which is based on FGD, the possibility to consider the pair/group meaning as a grammaticalized meaning of most Czech nouns was preferred to the possibility to treat it as a semantic feature of some of them (as a component of their lexical meaning). The latter possibility would have implied to split lexicon entries (at least) of the prototypical pair/group nouns into two entries, an entry with a common singular-plural opposition and an entry for cases in which the plural of the noun refers to pairs or groups (and behave, in fact, as pluralia tantum); the potential compatibility of the pair/group meaning with other nouns, though, would have remained unsolved. The broad coverage and the economy of the lexicon seem to be the main advantages that can be achieved when preferring the former solution in this case.

### 2.3. The pair/group meaning and the category of number

The pair/group meaning is closely connected with the grammatical number of nouns, though, we do not subsume it under this category; it is considered as a distinct one. The main reasons for this decision are, firstly, that the pair/group meaning is compatible both with singularity and plurality so that it cannot be considered as a third meaning of the category of number, and secondly, that the pair/group meaning is not subordinate to the meanings of the category of number so that it does not seem to be appropriate to consider it as a sub-value of singularity and plurality.

We thus worked with two oppositions in the theoretical description: the first opposition is the basic opposition of the category of number (i.e., sg vs. pl), the second one is constituted by the pair/group meaning (group) as opposed to the meaning of single entities (single). The combination sg-single (i.e. “one entity”) is expressed by singular forms of nouns, the other three combinations (sg-group “one group”, pl-single “more than one entity” and pl-group “more than one pair/group”)<sup>6</sup> by plural forms; see the annotation choices in Sect. 3 and process of matching the annotation choices to the grammatical values in Sect. 5.

The ambiguous plural form is disambiguated either by the numeral, which, however, co-occurs rather rarely in the data, or on the basis of context or knowledge of the world (thus in ex. (3) the combination sg-group is preferred whereas the same noun form in ex. (4) is interpreted as pl-group). This fact can be hardly used for an automatic identification of the particular meanings, thus we decided for a manual annotation of the pair/group meaning.

- (3) *Na rukou měl kožené rukavice.*  
 ‘He had leather **gloves** on his hands.’

---

<sup>6</sup>The combinations sg-group and pl-group are together referred to as the pair/group meaning in this paper.

- (4) *V obchodě nabízejí nejrůznější rukavice.*  
 ‘In the shop different **gloves** are sold.’

### 3. Manual annotation of the pair/group meaning in the PDT 2.0 data

#### 3.1. Selection of the data to be annotated

The aim of the annotation was to check whether the proposed pair/group meaning is distinctive enough in different contexts and how frequent it is in authentic language data. PDT 2.0 is a collection of Czech newspaper texts from 1990’s, to which morphological tagging and annotation at two syntactic layers was added: at the so-called analytical layer (layer reflecting the surface syntactic structure) and at the tectogrammatical layer (layer of the linguistic meaning of the sentence); at both of them the sentence is represented as a dependency tree with labeled nodes and edges.

As the pair/group meaning is expressed by formally unmarked plural forms, all plural forms of nouns are candidates for the manual disambiguation of the meanings studied here; i.e. 60 thousand plural noun forms out of all 833 thousand tokens for which tectogrammatical annotation is available. Nevertheless, since a rather low frequency of the pair/group meaning was expected on the background of a pilot annotation experiment,<sup>7</sup> only plural forms of those nouns were manually annotated for which the pair/group meaning is considered as prototypical, in order to make the annotation as efficient as possible. In the (open) list of prototypical pair/group nouns to be annotated, nouns were involved which co-occur with a set numeral in the PDT 2.0 and in the SYN2005 data, the list was further enriched using grammar books and theoretical studies on number in Czech<sup>8</sup> as well as linguistic introspection. The resulting list consists of 141 Czech nouns.<sup>9</sup>

*adidas* ‘adidas shoe’, *bačkora* ‘slipper’, *bačkorka/bačkůrka* ‘slipper.DIMIN’, *běžka* ‘cross-country ski’, *bok* ‘hip’, *bonbón* ‘bonbon’, *bota* ‘shoe’, *botaska* ‘botas shoe’, *botička* ‘shoe.DIMIN’, *botka* ‘shoe.DIMIN’, *brambor/brambora*

<sup>7</sup>In the pilot annotation 1,000 plural forms randomly selected from the SYN2005 corpus were involved, the pair/group meaning was preliminarily assigned with roughly 5 % of them. However, during the manual annotation of the PDT data, which is described in this paper, it turned out that the pair/group meaning is even much less frequent in the PDT data than in the pilot annotation. This fact might be connected with differences in the composition of the corpora (SYN2005 is a representative corpus of Czech, in PDT only newspapers are involved; see Sect. 4).

<sup>8</sup>Cf. esp. (Komárek et al., 1986), (Miko, 1962), and (Straková, 1960).

<sup>9</sup>The English translations capture the meaning of the listed Czech nouns; the formal characteristics of the English nouns thus do not correspond to those of the Czech ones in some cases (cf. the noun *těstovina*, which has both singular and plural forms in Czech, and its equivalent *pasta*). Some of the listed nouns are abbreviated from product names; well-known product names are included in the translation (cf. the noun *adidas* ‘adidas shoe’), if a specifically Czech product name is the source of the noun, its meaning is described without including the product name (*mišonka* ‘chocolate biscuit’). Diminutives are formed with special suffixes in Czech, they are marked with “.DIMIN” in the translations. If there exist formally different variants with the same meaning, they are introduced using the slash (*brambor/brambora* ‘potato’).

'potato', *brusle* 'skate', *chlup* 'hair', *chodidlo* 'sole', *cigareta* 'cigarette', *čtyřčce* 'quadruplet', *cvička* 'gym shoe', *datle* 'date', *dlaň* 'palm', *doklad* 'document', *dřeváček* 'clog.DIMIN', *dřevák* 'clog', *dvojče* 'twin', *fík* 'fig', *iniciála* 'initial', *kanada* 'working boot', *kapička* 'drop.DIMIN', *kapka* 'drop', *keks* 'cracker', *kel* 'tusk', *klíč* 'key', *klíček* 'key.DIMIN', *kolej* 'rail', *koleno* 'knee', *kolínko* 'knee.DIMIN', *končetina* 'limb', *kopačka* 'football boot', *kotník* 'ankle', *kozačka* 'boot', *křídlo* 'wing', *kroupa* 'barley', *kšanda* 'brace', *kulisa* 'scene', *kyčel* 'coxa', *lakýrka* 'patent shoe', *ledvina* 'kidney', *lék* 'medicine', *lentilka* 'chocolate candy', *lodička* 'pump', *loket* 'elbow', *lýtko* 'calf', *lyže* 'ski', *makaron* 'macaroni', *mandle* 'tonsil; almond', *mentolka* 'peppermint drop', *miňonka* 'chocolate biscuit', *mokasína* 'step-in shoe', *ňadro* 'breast', *náušnice* 'earring', *nehet* 'nail', *noha* 'foot, leg', *nozdra* 'nostril', *nožička* 'foot.DIMIN, leg.DIMIN', *nudle* 'noodle', *obočí* 'eyebrow', *očko* 'eye.DIMIN', *oko* 'eye', *oplatek/oplatka* 'wafer', *ořech* 'nut', *oříšek* 'nut.DIMIN', *osmerče* 'octuplet', *pantofle* 'slipper', *papuče* 'slipper', *parket/parketa* 'parquet', *paroh* 'horn', *partyzánka* 'cigarette', *pata* 'heel', *paterče* 'quintuplet', *piškot* 'sponge biscuit', *pistácie* 'pistachio', *plátěnka* 'canvas shoe', *plíce* 'lung', *podešev* 'sole', *podkolenka* 'knee sock', *ponožka* 'sock', *pouto* 'tie', *prarodič* 'grandparent', *prášek* 'pill', *prso* 'breast', *prst* 'finger', *punčocha* 'hose', *punčoška* 'hose.DIMIN', *rameno* 'shoulder', *řasa* 'eyelash', *ret* 'lip', *rodič* 'parent', *roh* 'horn', *rolnička* 'bell', *rozinka* 'raisin', *rtík* 'lip.DIMIN', *ručička* 'hand.DIMIN, arm.DIMIN', *ruka* 'hand, arm', *rukavice* 'glove', *sandál* 'sandal', *sardinka* 'sardine', *schod* 'stair', *schůdek* 'stair.DIMIN', *sedmerče* 'septuplet', *sirka* 'match', *sluchátko* 'earphone', *sourozenec* 'sibling', *sparta* 'cigarette', *stehno* 'thigh', *střevíc* 'shoe', *střevíček* 'shoe.DIMIN', *sušenka* 'biscuit', *šesterče* 'sextuplet', *škvarek/škvaarka* 'crackling', *šle* 'brace', *špageta* 'spaghetti', *teniska* 'gym shoe', *těstovina* 'pasta', *trojče* 'triplet', *tyčinka* 'bar', *ubrousek* 'napkin', *ucho* 'ear', *vlas* 'hair', *vločka* 'flake', *vráska* 'wrinkle', *zápalka* 'match', *zápěstí* 'wrist', *závora* 'barrier', *závorka* 'bracket', *žiletka* 'blade', *zoubek* 'tooth.DIMIN', *zub* 'tooth'

Only 67 out of the listed nouns were found in the PDT 2.0 data; for these noun lemmas there are 618 instances of plural forms in the data. More than a half of the 618 selected plural forms belong to five noun lemmas only (*oko* 'eye' 89, *rodič* 'parent' 87, *ruka* 'hand, arm' 81, *doklad* 'document' 35, *bota* 'shoe' 30; see the "coverage" in Table 4), 40 out of the 67 nouns had less than five instances of plural forms in the data. The plural forms to be annotated were extracted from the data together with a short context (both preceding and following) and divided into 31 html files. The annotators worked thus with a simple, linear text with highlighted plural forms followed by a drop-down list with five annotation choices, from which one should be chosen (see Fig. 1):<sup>10</sup>

- 1 - plurality,
- 2 - one pair/group,
- 3 - several pairs/groups,
- 4 - one pair/group or several pairs/groups,
- 5 - cannot be resolved.

<sup>10</sup>The default string "--" labeled as the sixth choice in the Tables was used by the annotators to indicate a mistake (for instance, if a singular form was involved in the annotation because of a mistake in the morphological tagging).

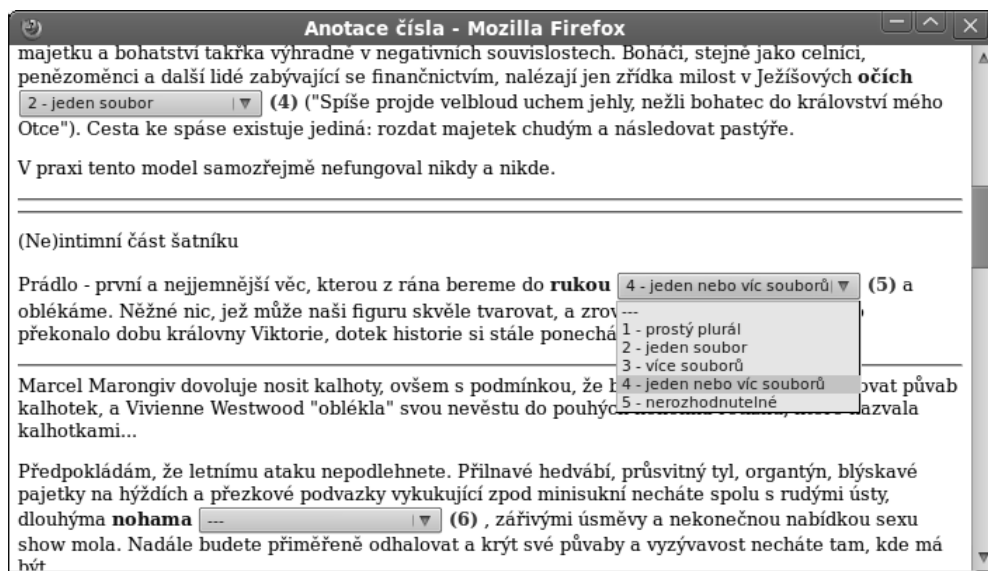


Figure 1. Screenshot of the html file to be annotated: linear text with highlighted instances followed by a drop-down list of annotation choices

### 3.2. Assigning the choices

All 31 files were annotated by two human annotators in parallel from October 2010 to January 2011, the annotation was preceded by a short training period. Both annotators are native Czech speakers; the language intuition of native speakers played a crucial role in the annotation process, several annotation “rules” formulated for problematic contexts are introduced in Section 3.3.

The first annotation choice, **1 - plurality**, was assigned to nouns denoting several single entities. The fact that single entities were referred to by the speaker was either obvious from the context (e.g. quantifier *několik* ‘several’ in ex. (5)) or could be inferred from the knowledge of the situation (cf. ex. (6)).

- (5) *Středem pozornosti kamer je nakrátko ostříhaná, křehká Dolores, která v červené čapce a s několika náušnicemi na pódiu vypadá jako kolébající se pirátka.*  
 ‘The close-cropped, flimsy Dolores is the center of the attention of the cameras, who in a red cap and with several earrings looks like a waddling pirate on the stage.’
- (6) *Sečíst pouhým okem stranickou příslušnost zvednutých rukou bylo ve dvousetčlenné Poslanecké sněmovně nemožné.*

'It was impossible to count up with the naked eye the party affiliation of the risen **hands** in the two-hundred member Chamber of Deputies.'

The second choice (**2 - one pair/group**) was the most frequently occurring choice in the annotation. It was assigned in basic contexts such as in ex. (7) (one human has one typical group of hairs on his head), but also in sentences like (8) in which the noun is used figuratively.

- (7) *Bydlí v Přadlácké ulici a má silnější, 178 cm vysokou postavu, hnědé krátké vlasy ...*  
'He lives in Přadlácká Street and is corpulent, 178 cm tall, has brown **hair** ...'
- (8) *Lidé od zaniklých pojišťoven se vrátí pod křídla VZP.*  
'People come back from the defunct insurance companies under the **wings** of the General Health Insurance Company.'

Unlike the previous choice, the annotators decided for the choice **3 - several pairs/groups** with less than 5 % of the annotated instances, mainly if the noun was accompanied by another noun in a close context which expresses the opposition of sg and pl regularly and was used in plural in the particular text. For instance, the noun *ruce* 'hands' in ex. (9) was assigned the choice 3 (which is in fact "plurality of pairs/groups") according to the plural form (plurality meaning) of the noun *hlavy* 'heads'.

- (9) *Šikovné ruce a hlavy rovněž nejsou tak vzácné. Thajské dívky dnes vyrábějí elektroniku světové úrovně.*  
'Even skillful **hands** and heads are not that rare. Today, Thai girls produce electronics of world-renowned quality.'

The choice **4 - one pair/group or several pairs/groups** has been proposed for cases in which the annotator preferred the pair/group meaning to the plurality meaning but was not certain which of the choices 2 or 3 is the right one. The annotators' uncertainty originated, on the one hand, from a lack of knowledge about the particular situation (ex. (10)) or, on the other, it was connected with the problem of expressing amounts in distributive contexts on the other. For instance, the plural *očíh* 'eyes' in ex. (11) can be interpreted both as several pairs because eyes of several people are denoted, and as one pair/group since each of the people should have glasses on his pair of eyes (cf. the noun *na očích* could be substituted by singular as well as plural form *na nose/nosech* 'on their nose/noses' in the particular Czech sentence; for the distributivity issue see Sect. 3.3). The choice one pair/group or several pairs/groups was the second most frequent choice in the annotation, nearly 25 % of the instances were assigned this value.

The annotation choice **5 - cannot be resolved** was used if there were neither linguistic features (context) nor extra-linguistic evidence that make the decision between the plurality meaning and the pair/group meaning possible (ex. (12)).



- (10) *Pro něho připravila firma Lotto speciální **kopačky**.*  
 ‘The Lotto company developed special **football boots** for him.’
- (11) *... aby lidé při sváření měli na **očích** ochranné brýle.*  
 ‘... so that people have protective glasses on their **eyes** during the welding.’
- (12) *... .. je to také odpověď na vzdělávací požadavky **rodičů**, žáků, ale i měnícího se trhu práce.*  
 ‘... it is an answer to educational requirements of the **parents**, pupils, but of the changing job market as well.’

### 3.3. Annotation of questionable cases: figurative usage, collectives, distributivity

Already during the short pre-annotation training, we came across many figurative contexts as well as phrasemes, titles etc. in which none of the proposed choices was intuitively preferred by the annotators. In order to achieve good annotation results even in these cases, we agreed on a rather general principle: the nouns should be interpreted in a possibly simple way. Thus for instance, the noun in ex. (8) mentioned above was treated as if we deal with a literal, non-figurative context (one pair/group), the nouns in the phrasemes in ex. (13) and (14) were assigned the same choice. A suggestion to exclude phraseological and figurative contexts from the annotation does not seem to be feasible in practice since in many cases the boundary between the literal and another kind of usage cannot be reasonably delimited.

- (13) *rozhovor z **očí do očí***  
 lit.: a talk from **eyes** to **eyes**  
 ‘a face-to-face talk’
- (14) *hnutí Na **vlastních nohou***  
 lit.: movement On own **feet**  
 ‘movement On one’s own two feet’

Another type of contexts discussed before the annotation started were sentences in which the noun to be annotated relates to a noun with a collective meaning, cf. the noun *oči* ‘eyes’ relates to the collective noun *posádka* ‘crew’, *uši* ‘ears’ to the noun *publikum* ‘audience’ etc. in ex. (15). According to the “rule” accepted by the annotators, such contexts were treated as if the whole group of persons referred to by the collective noun had just one pair of eyes, ears etc., thus the nouns marked in bold in ex. (15) were each assigned the second annotation choice (one pair/group). This rule might seem to be in conflict with an intuitive interpretation, since one can easily imagine the (exact) number of persons referred to by the noun *posádka* ‘crew’ in the particular context; however, taking into account that the nouns *publikum* ‘audience’ or *vláda* ‘government’ could be understood either as several individuals or as a body (this reading comes close to the ex. (8)) in the same context, the above rule proved to be useful for the annotators to keep consistency.

- (15) *před očima posádky, uším publika, v rukou vlády, do rukou státu*  
 ‘in front of the **eyes** of the crew, to the **ears** of the audience, in the **hands** of the government, into the **hands** of the state’

Distributivity, which is an issue extensively studied by formal semanticists and linguists, is addressed here just as affecting the annotation of the pair/group meaning (cf. (Dotlačil, 2010) who deals, among other languages, also with Czech). There is often a relation between a noun which was involved in the annotation and another noun of the sentence which refers to an amount of entities. If the entities denoted by the annotated noun relate to each of the entities referred to by the other noun, it is a case of distributivity; cf. ex. (11) where the noun *očích* ‘eyes’ denotes the distributed entities and the noun *lidé* ‘people’ the targets of the distribution.<sup>11</sup>

As for the nouns with a regular opposition of singular and plural in Czech, the distributed entities are expressed either by singular or plural when distributing one entity to each of the targets (ex. (16), (17)); at the tectogrammatical layer the nouns were assigned sg or pl according to their form, without taking into account their interchangeability).<sup>12</sup> Nouns we deal with are used always in plural if denoting distributed entities – when selecting one of the proposed choices, the question arose whether the plural should be interpreted as denoting one pair/group or several pairs/groups; the substitution test obviously does not help in such cases. The annotators decided with regard to the close context (ex. (18) and (19)); the noun *oči* ‘eyes’ in ex. (18) was assigned the choice several pairs/groups due to the plural form of the noun *nosy* ‘noses’, in ex. (19) the nouns *vrásky* ‘wrinkles’ and *oči* ‘eyes’ were both assigned one pair/group in accordance with the singular of the nouns *hlas* ‘voice’, *úsměv* ‘smile’ and *mluva* ‘speech’ in the particular sentence. In case there was no formal “clue” in the context, the choice one pair/group or several pairs/groups was assigned (ex. (20)). However, examples as (21), in which the choice was used twice (the noun *křídla* ‘wings’ was assigned the choice due to the lack of knowledge how many pairs of wings are concerned whereas the noun *nohy* ‘feet’ got this assignment due to the distributivity), has led us to the decision to distinguish the distributivity as a separate choice for the next annotation phase (see Sect. 4.2).

- (16) *Studenti kroutili **hlavou/hlavami**.*<sup>13</sup>  
 ‘Students shook their **head/heads**.’
- (17) *Studenti měli na **hlavě/hlavách** čapku/čapky.*  
 ‘Students had **hat/hats** on their **head/heads**.’

<sup>11</sup>The distributivity is, of course, not limited to the nouns expressing pairs/groups.

<sup>12</sup>Categories of nouns in distributive contexts can vary from language to language; cf. (Lashevskaja, 1999) for Russian and (Corbett, 2000) for other languages.

<sup>13</sup>In the examples (16) to (21) the nouns denoting the target of the distribution are underlined, the nouns with the pair/group meaning are marked in bold (as in the whole article), they express the distributed entities.

Choices by Annotator 1	Choices by Annotator 2						Total
	1	2	3	4	5	6	
1 - plurality	<b>115</b>	6	18	7	8	0	154
2 - one pair/group	5	<b>180</b>	1	16	5	0	207
3 - several pairs/groups	1	1	<b>22</b>	7	9	0	40
4 - one pair/group or several pairs/groups	2	27	14	<b>112</b>	15	0	170
5 - cannot be resolved	1	3	3	3	<b>35</b>	0	45
6 - - - -	0	1	0	1	0	<b>0</b>	2
Total	124	218	58	146	72	0	618

Table 1. Inter-annotator agreement in the annotation carried out on the PDT 2.0 data. The number of instances assigned each annotation choice by the first annotator are given in rows, the total number for each choice (the last column) is divided according to the choices by the second annotator, which are displayed in columns following the same principle (cf., 154 instances in total were assigned the choice 1 by the first annotator, in 115 out of them the second annotator assigned the same choice, in 6 of them the second annotator assigned the choice 2 etc.). Numbers of instances assigned the same choice by both annotators are marked in bold on the diagonal.

- (18) *Rozkvetlým městem chodí kýčející lidé s červenými nosy, oteklýma **očima**, plnýma slz, a z kapes jim vypadávají papírové kapesníky.*  
 ‘Through the flowering town, sneezing people with red noses are walking, with swollen **eyes**, full of tears, and paper tissues are falling out of their pockets.’
- (19) *Všude na světě to pánové dělají hlouběji posazeným hlasem, neprůstřelným úsměvem, pomalejší mluvou s dobře oddělovanými slovy, milým vějířkem **vrásek** kol **očí**.*  
 ‘All around the world men do it with help of a deeper-set voice, a bullet-proof smile, a slower speech with well-separated words, a nice fan of **wrinkles** around the **eyes**.’
- (20) *Divil jsem se, že mu to Američané povolili, když všechny zprávy procházely jejich **rukama**.*  
 ‘I was surprised that Americans allowed it to him, when all messages went through their **hands**.’
- (21) *... aby v dunění **křidel** dobyla vítězství Těch, kdo alespoň pomocí **nohou** uprchli tíže neplodného těla.*  
 ‘... so that in the rumble of **wings** she gains the victory of Those who, at least with help of the **feet**, escape the heaviness of the sterile body.’

### 3.4. Agreement analysis

The annotators agreed on 464 (75.1 %) out of 618 instances annotated, with a Kappa score of 0.67.<sup>14</sup> Another 64 instances were assigned either the choice 2 or 3 by the first annotator and the choice 4 by the second annotator, or vice versa. Thus, if having a less granular scale of annotation choices, an even higher agreement score might be expected. An overview of choices assigned by each of the annotators and the number of instances both annotators (dis)agreed on is given in Table 1.

After the parallel annotation had been finished, instances of disagreement were decided by a third annotator and the instances on which annotators agreed were revised in order to check the correctness and consistency of the annotation; the revised annotation is referred to as final annotation in the sequel.

With 69 of the 154 differently annotated instances, the choice of the first annotator was preferred, the choice of the second annotator was acknowledged to be the right one with 61 of the instances, the remaining 24 instances of disagreement were assigned a choice different from that of the first as well as the second annotator. Concerning the 464 instances which annotators agreed on, only three of them were changed by the third annotator during the revision.

In the final annotation, the annotation choice 2 was the most frequent one, see Table 2. As the choices 2, 3 and 4 are, in fact, particular meanings of the pair/group meaning, we can state that 414 plural forms were assigned the pair/group meaning in the presented annotation, i.e. in 67.0 % of the annotated instances (cf. the sum of choices 2, 3 and 4 in Table 2).

Further, we were interested in how frequent the pair/group meaning was with single noun lemmas which were involved in the annotation. For nouns *dvojče* 'twin', *pouto* 'tie', *ledvina* 'kidney', *vlas* 'hair', *kopačka* 'football boot', *ucho* 'ear', *lyže* 'ski', and *schod* 'stair', even all their instances were assigned one of the choices 2, 3 or 4. In Table 3 the percentage of the instances assigned the pair/group meaning among all instances of nouns with five or more plural occurrences in the PDT 2.0 data is specified. Table 4 gives a detailed overview of all annotation choices for each noun with five or more plural instances in the PDT 2.0 data, the numbers correspond to the final annotation; the inter-annotator agreement for each of these nouns is shown as well.

## 4. Pair/group meaning in the written vs. spoken data

### 4.1. Low frequency in the PDT 2.0 data

It is apparent from the analysis of the manual annotation that the pair/group meaning has a very low frequency in the PDT 2.0 data. We faced thus the question

---

<sup>14</sup>The Cohen's Kappa measure is used, which takes into account the effect of agreement by chance (Cohen, 1960).

Annotation choice	# of instances assigned	Percentage
1 - plurality	133	21.5 %
2 - one pair/group	230	37.2 %
3 - several pairs/groups	30	4.9 %
4 - one pair/group or several pairs/groups	154	24.9 %
5 - cannot be resolved	70	11.3 %
6 - - -	1	0.2 %
Total	618	100.0 %

Table 2. Annotation choices in the final annotation of the PDT 2.0 data

Noun lemma	# of plural forms			Noun lemma	# of plural forms		
	# of pl. forms with the pair/group meaning	Percentage	Percentage		# of pl. forms with the pair/group meaning	Percentage	Percentage
<i>dvojče</i> 'twin'	5	5	100.0 %	<i>noha</i> 'foot, leg'	20	17	85.0 %
<i>pouto</i> 'tie'	5	5	100.0 %	<i>kulisa</i> 'scene'	6	5	83.3 %
<i>ledvina</i> 'kidney'	7	7	100.0 %	<i>koleno</i> 'knee'	5	4	80.0 %
<i>vlas</i> 'hair'	11	11	100.0 %	<i>bota</i> 'shoe'	30	24	80.0 %
<i>kopačka</i> 'football shoe'	5	5	100.0 %	<i>klíč</i> 'key'	8	5	62.5 %
<i>ucho</i> 'ear'	9	9	100.0 %	<i>zub</i> 'tooth'	14	8	57.1 %
<i>lyže</i> 'ski'	13	13	100.0 %	<i>rodič</i> 'parent'	87	37	42.5 %
<i>schod</i> 'stair'	6	6	100.0 %	<i>křídlo</i> 'wing'	17	5	29.4 %
<i>ruka</i> 'hand, arm'	81	77	95.1 %	<i>doklad</i> 'document'	35	8	22.9 %
<i>prst</i> 'finger/toe'	10	9	90.0 %	<i>cigareta</i> 'cigarette'	17	3	17.6 %
<i>oko</i> 'eye'	89	80	89.9 %	<i>lék</i> 'medicine'	16	2	12.5 %
<i>rameno</i> 'shoulder'	9	8	88.9 %	<i>brambor</i> 'potato'	9	1	11.1 %
<i>rukavice</i> 'glove'	8	7	87.5 %	<i>těstovina</i> 'pasta'	7	0	0.0 %
<i>kolej</i> 'rail'	16	14	87.5 %	<b>Total</b>	<b>618</b>	<b>414</b>	<b>67.0 %</b>

Table 3. Noun lemmas with five or more plural instances in the PDT 2.0 data are arranged according to the percentage of instances assigned the pair/group meaning (i.e. the sum of the instances assigned the choices 2, 3 or 4) among all plural instances of these nouns in the final annotation.

Noun lemma	# of plural forms	Coverage	# of instances of agreement	Instances of agreement (%)						
					1 - plurality	2 - one pair/group	3 - several pairs/groups	4 - one pair/group or several pairs/groups	5 - cannot be resolved	6 - - - -
<i>oko</i> 'eye'	89	14.4%	67	75.3%	7	44	1	35	1	1
<i>rodič</i> 'parent'	87	28.5%	56	64.4%	1	28	0	9	49	0
<i>ruka</i> 'hand, arm'	81	41.6%	67	82.7%	4	35	2	40	0	0
<i>doklad</i> 'document'	35	47.2%	27	77.1%	26	5	0	3	1	0
<i>bota</i> 'shoe'	30	52.1%	18	60.0%	2	7	8	9	4	0
<i>noha</i> 'foot, leg'	20	55.3%	19	95.0%	3	12	1	4	0	0
<i>cigareta</i> 'cigarette'	17	58.1%	15	88.2%	14	1	2	0	0	0
<i>křídlo</i> 'wing'	17	60.8%	14	82.4%	12	3	0	2	0	0
<i>kolej</i> 'rail'	16	63.4%	8	50.0%	2	11	0	3	0	0
<i>lék</i> 'medicine'	16	66.0%	8	50.0%	8	0	2	0	6	0
<i>zub</i> 'tooth'	14	68.3%	8	57.1%	2	5	0	3	4	0
<i>lyže</i> 'ski'	13	70.4%	10	76.9%	0	2	0	11	0	0
<i>vlás</i> 'hair'	11	72.2%	9	81.8%	0	9	0	2	0	0
<i>prst</i> 'finger/toe'	10	73.8%	7	70.0%	1	8	0	1	0	0
<i>brambor</i> 'potato'	9	75.2%	8	88.9%	8	1	0	0	0	0
<i>rameno</i> 'shoulder'	9	76.7%	9	100.0%	1	6	0	2	0	0
<i>ucho</i> 'ear'	9	78.2%	7	77.8%	0	5	0	4	0	0
<i>klíč</i> 'key'	8	79.4%	5	62.5%	2	4	0	1	1	0
<i>rukavice</i> 'glove'	8	80.7%	6	75.0%	1	2	4	1	0	0
<i>ledvina</i> 'kidney'	7	81.9%	6	85.7%	0	4	0	3	0	0
<i>těstovina</i> 'pasta'	7	83.0%	6	85.7%	7	0	0	0	0	0
<i>kulisa</i> 'scene'	6	84.0%	3	50.0%	0	5	0	0	1	0
<i>schod</i> 'stair'	6	85.0%	4	66.7%	0	5	1	0	0	0
<i>dvojče</i> 'twin'	5	85.8%	5	100.0%	0	5	0	0	0	0
<i>koleno</i> 'knee'	5	86.6%	4	80.0%	1	2	0	2	0	0
<i>kopačka</i> 'football boot'	5	87.4%	4	80.0%	0	1	0	4	0	0
<i>pouto</i> 'tie'	5	88.2%	5	100.0%	0	1	0	4	0	0
Total	618	100.0%	464	75.1%	133	230	30	154	70	1

Table 4. Noun lemmas with five or more plural instances in the PDT 2.0 data, arranged according to their frequency. In the coverage column, the percentage of the instances of the first to the last of the listed nouns with respect to the total number of instances to be annotated is expressed. The inter-annotator agreement is specified for each noun both in number of instances and in percentage. In the right part of the Table, the number of the instances assigned the choices 1 to 5 (and 6) in the final annotation is shown for each noun.

Choices by Annotator 1	Choices by Annotator 2							Total
	1	2	3	4	5	6	7	
1 - plurality	<b>82</b>	8	1	2	1	3	0	97
2 - one pair/group	11	<b>350</b>	0	4	0	24	0	389
3 - several pairs/groups	0	0	<b>10</b>	0	2	7	0	19
4 - one pair/group or several pairs/groups	0	4	1	<b>4</b>	0	1	0	10
5 - cannot be resolved	0	0	0	0	<b>0</b>	1	0	1
6 - distributivity	1	10	1	0	4	<b>43</b>	0	59
7 - - -	3	0	0	0	0	0	<b>0</b>	3
Total	97	372	13	10	7	79	0	578

Table 5. Inter-annotator agreement in the annotation carried out on the spoken data of PDTSC. Numbers of instances assigned the same choice by both annotators are marked in bold on the diagonal.

of whether or not this meaning should be included in the forthcoming version of the treebank (PDT 3.0).

The 414 instances assigned the pair/group meaning (i.e. the choices 2, 3 or 4) during the annotation correspond to only 0.69 % of all 60,017 plural forms of nouns at the tectogrammatical layer. However, if we compared the number of instances of the pair/group meaning to the frequency of other attributes annotated at the tectogrammatical layer, namely to the frequency of functor values (i.e. dependency relations, semantic roles), 14 functors (out of 67) do 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”. There are also several grammatemes whose values are less frequent than the pair/group meaning (for instance, only 375 tectogrammatical nodes are assigned the value *imp* in the grammateme of verbal mood corresponding to the imperative mood of verbs).

#### 4.2. A higher frequency in the spoken data

Taking into account the fact that, as already mentioned, in PDT 2.0 only written newspaper texts are involved, we were wondering whether the frequency of the pair/group meaning would be different (higher) in spoken data or in written data from other genres.

After the manual annotation of the PDT 2.0 data had been finished, a manual annotation was carried out on the data from the Prague Dependency Treebank of Spoken Czech (PDTSC), which is currently built at the Institute of Formal and Applied Linguistics at Charles University in Prague (Hajič et al., 2008).<sup>15</sup> The instances to be

<sup>15</sup>PDTSC is the Czech part of the Prague Dependency Treebank of Spoken Language (PDTSL, <http://ufal.mff.cuni.cz/pdtsl>).

Annotation choice	# of instances assigned	Percentage
1 - plurality	106	18.3%
2 - one pair/group	380	65.7%
3 - several pairs/groups	12	2.1%
4 - one pair/group or several pairs/groups	7	1.2%
5 - cannot be resolved	5	0.9%
6 - distributivity	67	11.6%
7 - - -	1	0.2%
Total	578	100.0%

*Table 6. Annotation choices in the final annotation of the data from PDTSC*

annotated were selected from the tectogrammatically annotated data of PDTSC (316 thousand tokens for which tectogrammatical annotation was available at that moment) using the same procedure as described in Sect. 3.1. The annotation was carried out by the same two annotators. The list of annotation choices was enriched with the choice distributivity for nouns in distributive contexts (see Sect. 3.3), so that the choice one pair/group or several pairs/groups was used only in clear contexts exemplified by the ex. (10).

The annotators agreed on 489 out of 578 annotated plural nouns, i.e. on 84.6 % of the instances, Kappa score 0.71. The choices assigned by the annotators are compared in Table 5, number of instances assigned the particular choices in the final annotation are listed in Table 6.

For the spoken data, a significantly higher inter-annotator agreement was achieved than for the written data. The percentage of the instances assigned the pair/group meaning among all annotated instances was also higher in the data from PDTSC than from PDT 2.0; see the sum of choices 2, 3 and 4 in Table 2 (67.0 %) vs. choices 2, 3, 4 and 6 in Table 6 (80.6 %; the new choice distributivity is another particular value of the pair/group meaning). This difference is related, among other facts, for instance to a higher frequency of every-day contexts and a lower frequency of figuratively used nouns and phrasemes in the spoken than in the written data.

The hypothesis that the relatively low frequency of the pair/group meaning of the PDT 2.0 data has a relation to the type of texts involved in the treebank is further supported by a comparison with three large corpora of written Czech texts, namely with balanced corpora which were built in the Czech National Corpus;<sup>16</sup> see Table 7. In the table, several corpora are compared only as to the number of plural forms of nouns from the working list of pair/group nouns; the manual annotation of the pair/group meaning was not provided for all the corpora. Nevertheless, we conclude on this background that the involvement of the annotation of the pair/group meaning in the PDT 3.0 is worth the effort.

<sup>16</sup><http://ucnk.ff.cuni.cz>



Corpus	# of plural forms of nouns from the working list	Size of the corpus (# of tokens)	# of noun forms	# of plural forms of nouns
PDT 2.0	618	833,195 0.07 %	256,271 0.24 %	60,017 1.03 %
PDTSC	578	316,086 0.18 %	48,976 1.18 %	12,104 4.78 %
SYN2000	161,004	120,908,724 0.13 %	32,479,355 0.50 %	7,712,904 2.09 %
SYN2005	271,949	122,419,382 0.22 %	31,315,440 0.87 %	7,440,382 3.66 %
SYN2010	273,680	121,667,413 0.22 %	29,808,857 0.92 %	7,225,687 3.79 %

*Table 7. Number of plural instances of the nouns for which the pair/group meaning is supposed to be prototypical (see the working list) in the PDT 2.0 and PDTSC data compared with three sub-corpora of the Czech National Corpus. The percentage stated below the numbers of instances in each row is the percentage of the plural forms of the nouns from the working list among all tokens, all noun forms and all plural forms of nouns, respectively, in each particular corpus.*

## 5. Matching the annotation on the data

### 5.1. Inserting the manual annotation into the data

As explained in Section 2, the pair/group meaning is treated as a grammaticalized meaning constituting a new grammatical category of Czech nouns, which is closely related to the category of noun number. In FGD as well as in the annotation scenarios of PDT 2.0 and PDT 3.0, which are based on this theoretical framework, grammatical meanings are captured within the so-called *grammatemes*, which are attributes of nodes of the tectogrammatical tree. *Grammatemes* correspond to morphological categories, such as number with nouns, degree of comparison with adjectives or tense with verbs.<sup>17</sup>

For the purpose of including the pair/group meaning into the tectogrammatical annotation of PDT 3.0, a new *grammateme* *typgroup* was added to the existing set of 15 *grammatemes* used in PDT 2.0 (Mikulová et al., 2006). For the *typgroup* *grammateme*, two values were defined: *group* for the pair/group meaning and *single* for the meaning of single entities. To be able to represent all the semantic nuances distinguished in the manual annotation (choices 1 to 5 in Sect. 3) in the treebank data, the values of the *grammateme* *typgroup* must be combined with the values of the *grammateme* number. For each of the both *grammatemes*, a third value (*nr*, “not recognized”) is used.

The annotation choices 1 to 5 were matched to the values of the *grammatemes* number and *typgroup* as follows. The annotation choice is given first, the arrow is followed by the values of the number and *typgroup* *grammatemes*, respectively:

- 1 - plurality → number=*pl*, *typgroup*=*single*
- 2 - one pair/group → number=**sg**,<sup>18</sup> *typgroup*=*group*
- 3 - several pairs/groups → number=*pl*, *typgroup*=*group*
- 4 - one pair/group or several pairs/groups → number=**nr**, *typgroup*=*group*
- 5 - cannot be resolved → number=**nr**, *typgroup*=*nr*

### 5.2. Automatic annotation of the pair/group meaning with remaining nouns

Since, according to our proposal, the pair/group meaning concerns potentially all Czech nouns, nouns which were involved neither in the list and, thus, nor in the manual annotation, were assigned a value of the *typgroup* *grammateme* fully automatically. A simple, two-step “algorithm” was provided for the automatic annotation: in the first step, nouns accompanied with a set numeral *jedny* ‘one pair/group’ (pluralia tantum excepted) were assigned the value *group* of the *typgroup* *grammateme* and the

<sup>17</sup>Unlike the mentioned categories, there are no *grammatemes*, e.g., for number of adjectives or case of nouns since these categories are imposed by agreement or government, respectively.

<sup>18</sup>Those number values are marked in bold which were changed from the *pl* value (as available in the PDT 2.0 annotation) to the marked value, influenced by the annotation of the pair/group meaning.

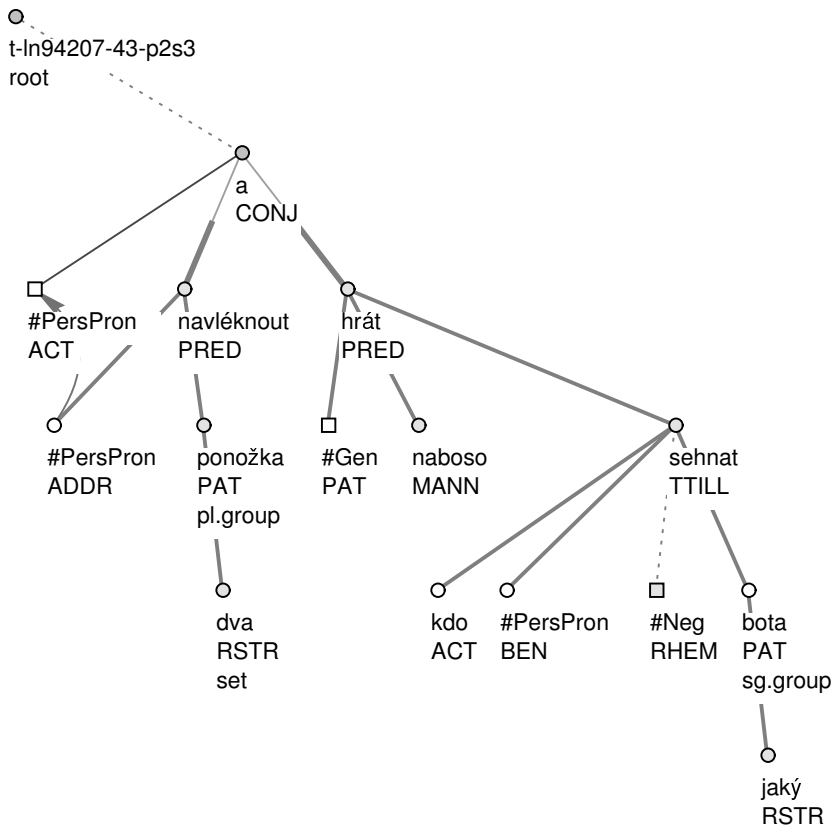


Figure 2. Tectogrammatical tree of the sentence “Navlékla bych si dvoje ponožky a hrála bych naboso, dokud by mi někdo nesehnal nějaké boty.” ‘I would put on two pairs of socks and would play barefooted until somebody would get some shoes for me.’ For each node the tectogrammatical lemma, the functor and the values of the grammatemes number, tygroup and numertype are given; the numertype (value set) is assigned only to the node with the lemma “dva” ‘two’, which represents the set numeral “dvoje” ‘two pairs/sets’.

value of the number grammateme was changed to sg in this connection; if the noun collocated with a set numeral of a higher numeric value (*dvoje* ‘two pairs/groups’, *troje* ‘three pairs/groups’ etc.), the value group was filled in the grammateme *typgroup* whereas the number grammateme remained unchanged (i.e. pl). Secondly, all the other nouns were assigned the value single in the *typgroup* grammateme, the value of the number grammateme was not changed in these cases, compared to the PDT 2.0 data. A sample tectogrammatical tree with nodes assigned the number and *typgroup* values is displayed in Fig. 2.

## 6. Conclusions

The main focus of the present paper has been laid on the manual assignment of the pair/group meaning with selected Czech nouns. With regard to the fact that the pair/group meaning is a very semantic issue, which is complicated with a strong ambiguity and has been studied only recently for Czech, the achieved inter-annotator agreement is rather satisfactory. The manual annotation was completed with the automatic assignment of *typgroup* values to the nouns which were not involved in the manual part. In PDT 3.0, thus, all nouns will be assigned the pair/group meaning.

We have in mind that there are several other issues in the domain studied here that are open for further investigation, for instance, (a) systematic study of the numerals from the point of view of their form and function with regard to their compatibility with the different types of nouns, (b) consequences of (a) for the deep-lexical representation of the different types of numerals in lexicon, outcoming from the preliminary solution given in (Razímová and Žabokrtský, 2006), (c) consideration about possibilities and limits of the semi-automatic annotation of quantified noun phrases as to the grammatememes number and *typgroup* and its implementation.

## Acknowledgments

The work reported on in the present paper has been supported by the grant projects GA ČR P406/2010/0875 and MŠMT ČR LC536. We would like to thank to Eva Fernandez de Jesus and Hana Hanová who carried out the annotation described in the paper.

## Bibliography

- Cohen, Jacob. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46, 1960.
- Corbett, Greville G. *Number*. Cambridge University Press, Cambridge, 2000.
- Dotlačil, Jakub. *Anaphora and Distributivity*. LOT, Utrecht, 2010.
- Hajič, Jan, Silvie Cinková, Marie Mikulová, Petr Pajas, Jan Ptáček, Josef Toman, and Zdeňka Uřešová. PDTSL: An annotated resource for speech reconstruction. In *Proceedings of the 2008 IEEE Workshop on Spoken Language Technology*, pages 93–96, Goa, India, 2008. IEEE.

- Hajič, Jan, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, and Magda Ševčíková Razímová. *Prague Dependency Treebank 2.0*. Linguistic Data Consortium, Philadelphia, 2006.
- Komárek, Miroslav, Jan Kořenský, Jan Petr, and Jarmila Veselková et al. *Mluvnice češtiny 2*. Academia, Praha, 1986.
- Lashevskaja, Olga N. Chislovoe oformlenie predmetnyh suschestvitel'nyh v distributivnom kontekste. *Nauchnaja i tekhnicheskaja informacija. Ser. 2. Informacionnyje processy i sistemy*, 11: 30–36, 1999.
- Miko, František. *Rod, číslo a pád podstatných mien*. Vydavateľstvo Slovenskej akadémie vied, Bratislava, 1962.
- Mikulová, Marie, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Uřešová, Kateřina Veselá, and Zdeněk Žabokrtský. Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual. Technical Report 30, ÚFAL MFF UK, Prague, Czech Rep., 2006.
- Panevová, Jarmila and Magda Ševčíková. Jak se počítají substantiva v češtině: poznámky ke kategorii čísla. *Slovo a slovesnost*, 72:163–176, 2011.
- Razímová, Magda and Zdeněk Žabokrtský. Annotation of Grammatemes in the Prague Dependency Treebank 2.0. In *Proceedings of the LREC 2006 Workshop on Annotation Science*, pages 12–19, 2006.
- Sgall, Petr. *Generativní popis jazyka a česká deklinace*. Academia, Praha, 1967.
- Sgall, Petr, Eva Hajičová, and Jarmila Panevová. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. D. Reidel Publishing Company, Dordrecht, 1986.
- Straková, Vlasta. Ke kategorii čísla. In *Rusko-české studie*, pages 75–85, 1960.

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