Radek Čech*, Jiří Milička, Ján Mačutek, Michaela Koščová, Markéta Lopatková

Quantitative Analysis of Syntactic Dependency in Czech

Abstract: The article presents a quantitative analysis of some syntactic dependency properties in Czech. A dependency frame is introduced as a linguistic unit and its characteristics are investigated. In particular, a ranked frequencies of dependency frames are observed and modelled and a relationship between particular syntactic functions and the number of dependency frames is examined. For the analysis, the Czech Universal Dependency Treebank is used.

Keywords: syntactic dependency; dependency frame; syntactic function

1 Introduction

A hierarchical structure of a sentence can be expressed by the dependency grammar formalism (Mel'čuk 1998; Hudson 2007). This formalism describes the structure of a sentence in a form of a tree graph: nodes of the graph represent words, while links between nodes represent syntactic relationships between words. Within the approach, there is a syntactic function assigned to each word in a sentence, e.g., predicate, subject, object, determiner, etc. The dependency formalism allows us to gain an insight into relationships among sentence elements and it represents widely accepted formalism for description of syntactic properties of language. In this study, we have decided to adopt this approach for an analysis of more general properties of syntax; namely, we focus on the frequency distribution of so called dependency frames (for details, see Section 2) as well as on the relationship between the frequency of syntactic functions (subject, predicate, etc.) and the number of dependency frames of particular units. This kind of analysis is based on an assumption that a regular frequency

Radek Čech, University of Ostrava, Ostrava, Czech Republic, cechradek@gmail.com
Jiří Milička, Charles University, Prague, Czech Republic
Ján Mačutek, Comenius University in Bratislava, Bratislava, Slovakia
Michaela Koščová, Comenius University in Bratislava, Bratislava, Slovakia
Markéta Lopatková, Charles University, Prague, Czech Republic

https://doi.org/10.1515/9783110573565-003

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

distribution of language units can be interpreted as a result of very general facets of human language behavior. Specifically, the regular distribution is explained as a consequence of the least effort principle (Zipf 1949) or as an outcome of a diversification process in the language (Altmann 2005). Moreover, the regular distribution as well as the relationship between the frequencies of given units play a fundamental role in the synergetic model of language, which makes it possible to describe and, most importantly, to explain mutual interrelations among various language properties (Köhler 1986, 2005, 2012). Based on an expectation that relationships between syntactic functions are ruled by the same general mechanisms as other language properties, we set up the following hypotheses:

1) there is a regular frequency distribution of dependency frames in general in a language;
2) there is a regular frequency distribution of dependency frames for each syntactic function; differences among the distributions of individual syntactic functions are caused by their specific syntactic properties; differences are manifested by different models or different parameter values in the same model;
3) the more frequent the syntactic function, the more dependency frames it has.

This study represents a further step in the endeavor to apply quantitative methods in the dependency syntax analysis (e.g., Liu et al. 2017; Liu 2009; Mačutek et al. 2017; Čech et al. 2017).

The article is organized as follows. Section 2 introduces main characteristics of dependency frame. Section 3 describes the language material and the methodology. Section 4 discusses the results of the study and Section 5 concludes the article.

2 Dependency frame

According to the dependency grammar formalism, the (surface) syntactic structure of sentence (1) can be expressed by the tree graph displayed in Fig. 1.

(1) Christiane gave you a good answer

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

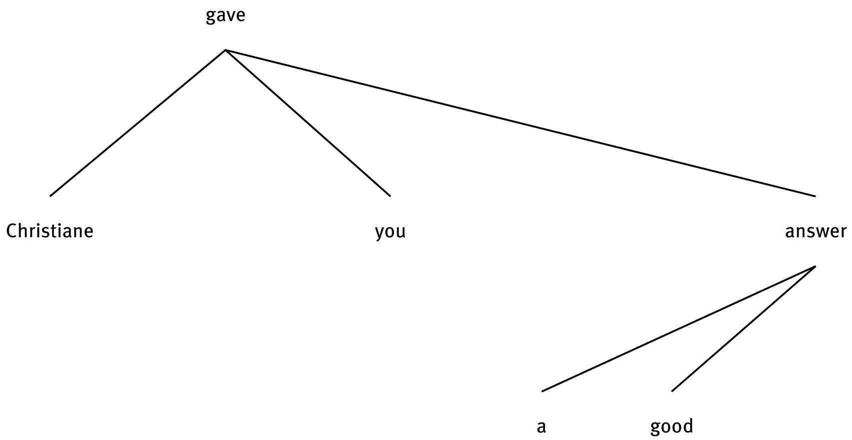

Fig. 1: The (surface) syntactic structure of sentence (1).

Following the approach presented by Čech et al. (2010), we set up a dependency frame (hereafter DF) as a basic unit of our analysis. Specifically, the DF is defined as a set of syntactic functions assigned to all words directly dependent on a given node in the tree. Particular syntactic functions are determined in accordance with the annotation used in the Universal Dependencies project (UDs)[1] (Zeman 2015; Nivre et al. 2016). The syntactic annotation within the UDs is based on the so called Stanford dependencies (de Marneffe et al. 2006, 2008, 2014). For illustration, the UDs annotation of sentence (1) is presented in Fig. 2. Syntactic functions (within the UDs project, they are called syntactic relations) are assigned to links between pair of words; they display the syntactic function of a word to which an arrow points. Thus, *Christiane* represents the nominal subject (nsubj), *you* represents the indirect object (iobj), *a* represents the determiner (det), etc.

There are two frame evoking words in sentence (1), *gave* and *answer,* and thus two DFs are identified there. In particular, the predicate *gave* has the frame:

[nsubj; iobj; obj]

1 http://universaldependncies.org/

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

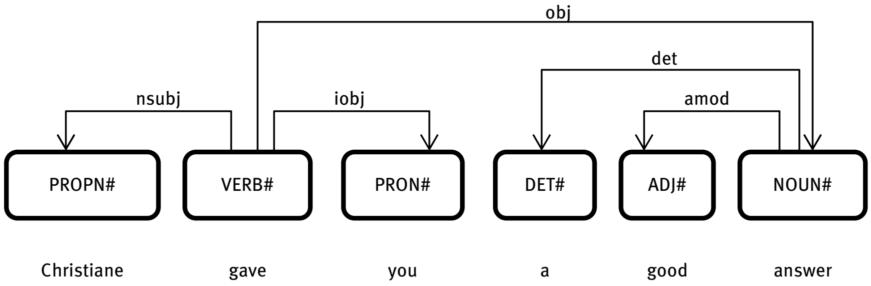

Fig. 2: The annotation of sentence (1) based on the principles used in the UDs project.

as the words *Christiane*, *you* and *answer* are directly dependent on the word *gave* (punctuation is disregarded here, see below); similarly, the object *answer* has the frame:

[det; amod]

as the words *a* and *good* are annotated as directly dependent on the word *answer* according to the UDs principles. All other words in sentence (1) have no directly dependent words, thus, no other DF can be determined there. In general, the notion of the frame evoking word (hereafter FEW) denotes a word which is a parent of DF elements in a syntactic tree; in other words, FEWs are all non-terminal elements of syntactic tree.

It should be noticed that word order is not taken into account as we are focusing on Czech, the language with high degree of word order freedom, where syntactic functions are represented by other means (esp. morphology) than the word order.

3 Language material and methodology

The Czech Universal Dependency Treebank, based on the Prague Dependency Treebank 3.0 (Bejček et al. 2013), is used in this study. The treebank consists of 87,913 sentences and about 1.5 million words/tokens. Its domain is mainly newswire, reaching also to business and popular scientific articles from the 1990s.[2]

2 see https://github.com/UniversalDependencies/UD_Czech/blob/dev/README.md

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

The procedure briefly sketched in Section 2 was applied to all sentences in the corpus. We used syntactic functions for determining DFs presented in Tab. 1.

In the analysis, we employed 21 syntactic functions (out of the total of 37 functions used in the UDs annotation scheme). We have omitted technical functions (such as punctuation, unspecified dependency, etc.) and linguistically not well-established functions (such as flat multiword expression, orphan, etc.). In sum, we have followed the UDs approach according to which syntactic functions presented in Tab. 1 represent dependency relations in their narrow sense (cf. http://universaldependencies.org/u/dep/index.html).

Tab. 1: The list of syntactic functions used for the determination of DF.

Syntactic function (relation)	**Abbreviation**
nominal subject	nsubj
object	obj
indirect object	iobj
oblique nominal	obl
vocative	vocative
nominal modifier	nmod
appositional modifier	appos
numeric modifier	nummod
clausal subject	csubj
clausal complement	ccomp
open clausal complement	xcomp
adverbial clause modifier	advcl
adverbial modifier	advmod
discourse element	discourse
clausal modifier of noun (adjectival clause)	acl
adverbial modifier	amod
copula	cop
determiner	det
case marking	case
auxiliary	aux
marker	mark

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

The UDs annotation was also used to determine syntactic functions of FEWs. However, it seems reasonable to slightly modify the original annotation scheme for predicates since they are not explicitly annotated in the UDs. Namely, as for main clauses, we assign the predicate function (pred) to the root node of the tree represented by a verb, see sentence (2), or to the non-verb root node on which a word with the "auxiliary" (AUX) POS tag is directly dependent, see sentence (3). Thus, in sentence (2) we determine the verb *comes* as the predicate, see Figure 3,

(2) From the AP comes this story

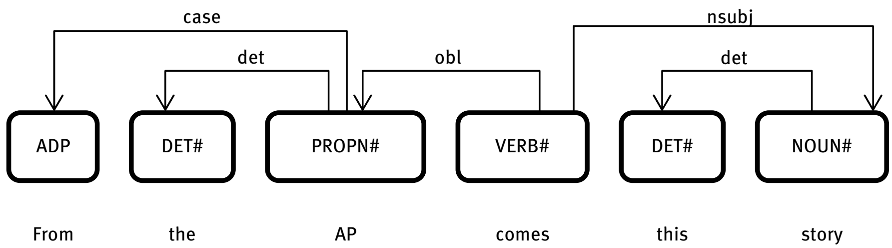

Fig. 3: The annotation of the sentence (2) based on the annotation used in the UDs project.

and in sentence (3) *young* is determined as the predicate (the auxiliary verb *was* is directly dependent on this word), see Figure 4.

(3) I was very young

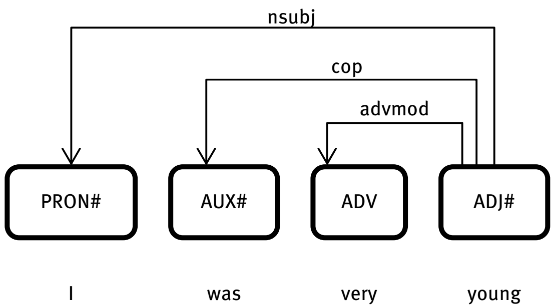

Fig. 4: The annotation of sentence (3) based on the annotation used in the UDs project.

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

Further, according to the UDs principles, predicates of dependent clauses are annotated in accordance with the function of the whole clause, as e.g. the clausal subject for the predicate of a subject clause, the adverbial clause modifier for the predicate of an adverbial clause (this is used also for determining DFs, see Tab. 1). In these cases, we assign the predicate function (pred) to words with syntactic functions csubj, ccomp, xcomp, advcl, and acl (which correspond to the root of clauses, in fact) if they are represented by verbs (as in sentence (4)), or to non-verb nodes on which a word with the “auxiliary” (AUX) POS tag is directly dependent. Consequently, in sentence (4) we determine the word *prove* (i.e., the verbal root of the sentence) as the predicate of the main clause, and *lost* (the verb with assigned syntactic function ‘accom’) as the predicate of the dependent clause, see Fig. 5.

(4) Today's incident proves that Sharon has lost his patience

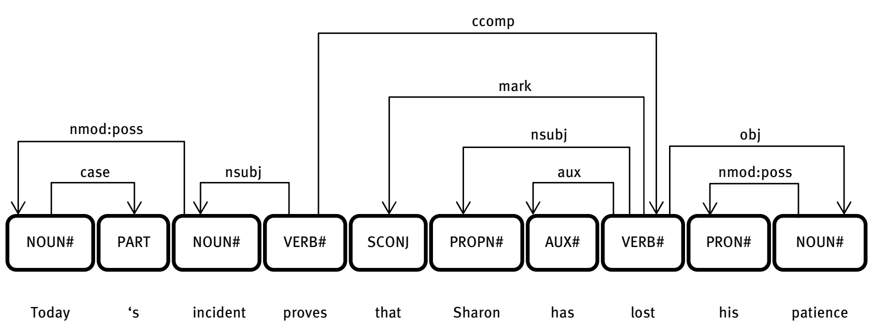

Fig. 5: The annotation of sentence (4) based on the annotation used in the UDs project.

All frequencies and numbers of occurrences were modelled by function

$$y = ax^b \tag{1}$$

There are several reasons for the choice of this particular mathematical model. First, it fits the data sufficiently well, as values of the determination coefficient are mostly greater than 0.9 (which is usually accepted as one of the criteria for a satisfactory fit in mathematical modelling of linguistic data), or at least not much lower (cf. Mačutek and Wimmer 2013 for a discussion of several goodness-of-fit measures applied in quantitative linguistics). Second, this model is very simple, nevertheless it reveals some differences among syntactic functions (see

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

Section 4). Finally, it is a special case of a very general mathematical formula expressing many language laws, which was presented by Wimmer and Altmann (2005). We thus remain within the general theoretical framework of quantitative linguistics, which makes it possible to investigate interrelations between syntax and other linguistic levels in future.

4 Results

The first hypothesis states that there should be a regular frequency distribution of DFs in the corpus. We have operationalized the hypothesis in two ways. First, all particular DFs and their frequencies were determined. Applying this procedure, we obtain 1,214 different DFs. The ten most frequent DFs are presented for illustration in Tab. 2 - not surprisingly, the most frequent DFs consists of one dependent word; the most frequent DFs with two dependents is represented by rank 4.[3]

Tab. 2: The ten most frequent DFs in the corpus.

Rank	Dependency frame	Frequency
1	nmod	5063
2	case	4569
3	amod	4454
4	case nmod	2262
5	amod case	2176
6	amod nmod	1613
7	obj	1117
8	det	1102
9	advmod	943
10	obl	864

Fitting function (1) to the data, we obtain parameter values a = 6375.13, b = −0.8737, with the determination coefficient R^2 = 0.897. Thus, our data corrobo-

3 All the data used for the experiments can be found at:
http://www.cechradek.cz/data/Cech_etal_Q_analysis_of_synt_dependency_results.zip

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

rate the first hypothesis as the value of determination coefficient is satisfactory. This result means that DF is a language unit which displays the rank-frequency distribution similar to distributions of the majority of well-established language units, such as words, lemmas, syllables, etc.

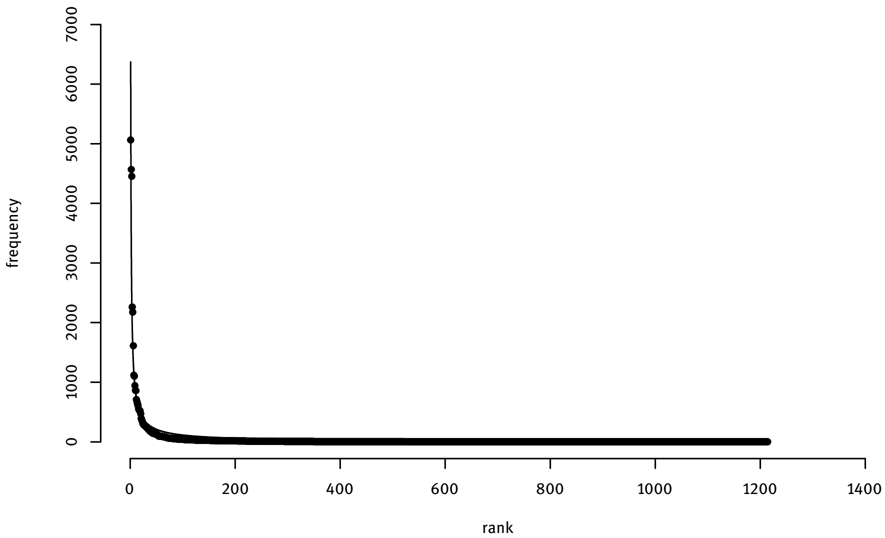

Fig. 6: Ranked frequencies of DFs in the corpus fitted with function (1).

As for the second way of the operationalization of the first hypothesis, we counted the number of all occurrences of DFs (i.e., their frequency in the data) for each chosen syntactic function as well as the number of unique DFs (i.e., the number of their types in the data). The results are presented in Tab. 3 and Fig.7 and Tab. 4 and Fig. 8, respectively.

Tab. 3: Frequency of all occurrences of DFs for each syntactic function of FEW in the data.

Rank	Syntactic function of FEW	Frequency of all DFs
1	pred	13973
2	nmod	10817
3	obl	8400
4	nsubj	5258

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

Rank	Syntactic function of FEW	Frequency of all DFs
5	obj	4451
6	amod	1361
7	appos	600
8	advmod	356
9	iobj	218
10	nummod	123
11	det	74
12	mark	7
13	vocative	7
14	discourse	1

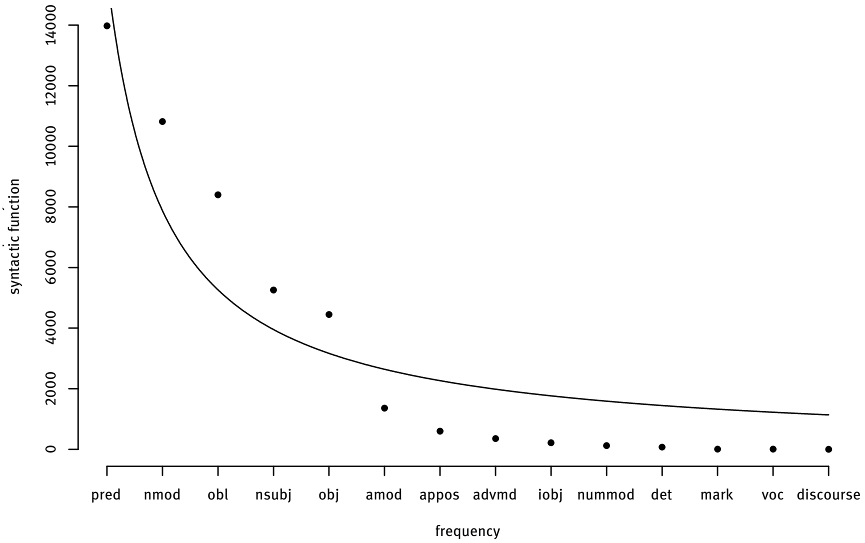

Fig. 7: Frequency of all occurrences of DFs for each syntactic function of FEW. The line represents function (1) with parameters $a = 15687.85$, $b = -0.9941$.

Fitting function (1) to the data from Tab. 3, we get following results: a = 15687.85, $b = -0.9941$, $R^2 = 0.849$, which is considered as acceptable result and thus the hypothesis is not rejected. However, the value of $R^2 < 0.9$ can/must be interpreted as a sign of an existence of fluctuations, which can be caused by

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

various reasons (e.g., the way of annotation, the character of the corpus) or/and as a sign of an unsuitability of the model used for the analysis. It is needless to say that only further research can reveal reasons for this phenomenon.

Fitting function (1) to the data from Tab. 4, we get $a = 929.42$, $b = -1.8313$, with $R^2 = 0.959$. In this case, we obtain very satisfactory fit in terms of the determination coefficient, thus, the hypothesis can be tentatively considered as corroborated, too.

The analysis of results for both frequency distributions (i.e., all DFs occurrences and number of unique DFs, i.e. types), a specific position of predicate is revealed. It can be explained as a consequence of its central role in a clause structure and its specific position in the syntactic tree - it is always the root of the tree (or the root of the subtree representing a dependent clause) and, consequently, it is not influenced by hierarchically higher syntactic elements.

Tab. 4: Number of unique DFs (DF types) for each syntactic function of FEWs.

Rank	Syntactic function of FEW	Number of unique DFs (DF types)
1	pred	947
2	nmod	146
3	obl	138
4	obj	136
5	nsubj	132
6	appos	112
7	amod	54
8	iobj	29
9	advmod	18
10	nummod	13
11	det	8
12	mark	2
13	vocative	2
14	discourse	1

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

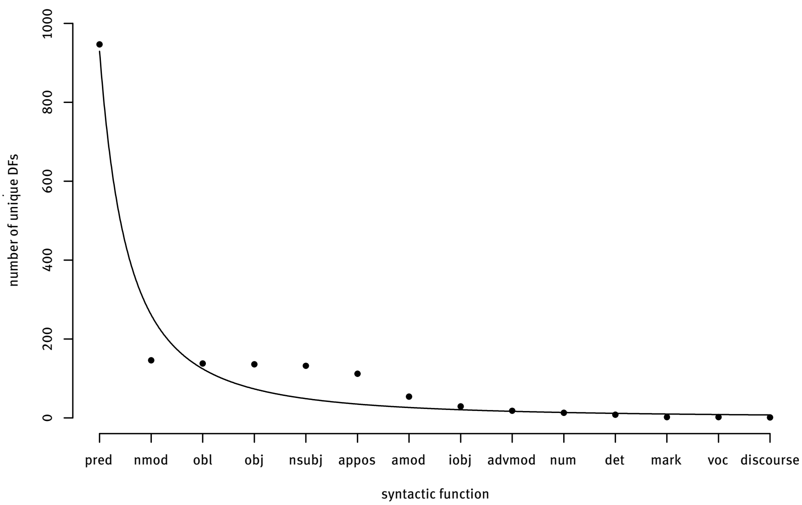

Fig. 8: Number of unique DFs (DF types) for each syntactic function. The line represents function (1) with parameters: $a = 929.42$, $b = -1.8313$.

Tab. 5: Results of fitting function (1) to the data.

Syntactic function	Parameter *a*	Parameter *b*	R^2
pred	1154.18	−0.7603	0.921
nmod	2533.42	−0.9551	0.872
iobj	71.24	−1.0954	0.945
obj	1341.97	−1.1111	0.942
appos	151.25	−1.1513	0.996
nsubj	1805.06	−1.2124	0.966
obl	2903.76	−1.2211	0.969
amod	558.25	−1.3165	0.955
advmod	161.20	−1.3288	0.952
nummod	59.27	−1.4099	0.979
det	41.37	−1.5830	0.993

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

To get a deeper insight into the analysed phenomena, we observed frequency distributions of DFs for each syntactic function of FEWs as well. We hypothesize that there is a regular frequency distribution of dependency frames for each syntactic function of FEWs (c.f. hypothesis (2) above). To test this hypothesis, we counted ranked frequencies of DFs for each syntactic function separately and then fit function (1) to the data. The results, presented in Tab. 5, show regular ranked frequencies in all cases - the determination coefficient lies in the interval <0.872, 0.996>; this means that hypothesis (2) is not rejected. Further, the coefficient *b* differs with regard to particular syntactic functions. For illustration, the ranked frequencies of syntactic functions with extreme values of *b* are presented in Figures 9 and 10.

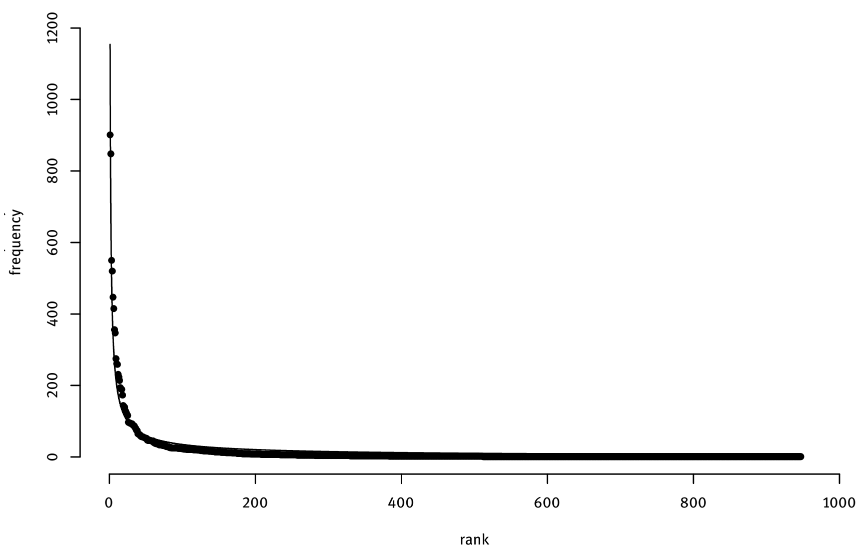

Fig. 9: Ranked frequencies of predicates' DFs (with the highest value of $b = -0.7603$).

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

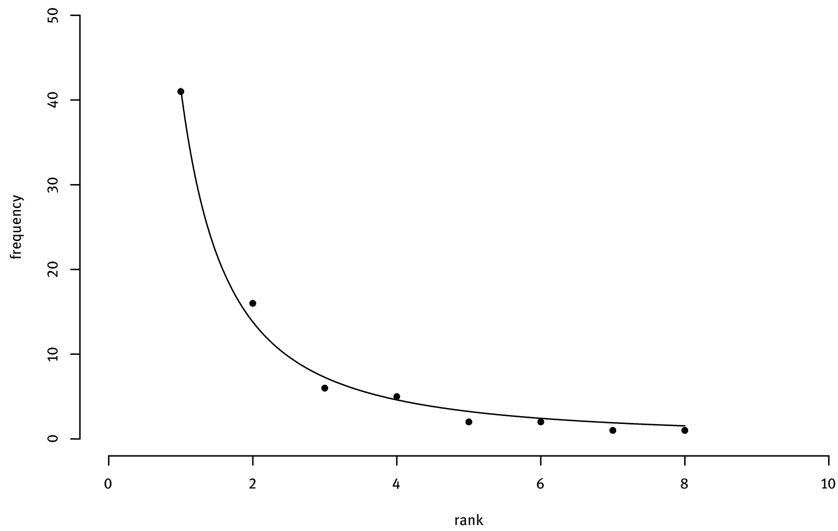

Fig. 10: Ranked frequencies of determinators' DFs (with the lowest value of $b = -1.5830$).

According to the third hypothesis, there should be a relationship between the frequency of FEW with the given syntactic function and the number of unique DFs (DF types) for the given syntactic function. Specifically, we hypothesize: the more frequent the syntactic function, the more unique DFs it has. The hypothesis was tested on the data which are presented in Tab. 6 and Fig. 11.

Tab. 6: Frequencies of FEW with the given syntactic function and of unique DFs (DF types) for the given syntactic function.

Syntactic function	Parameter *a*	Parameter *b*	R^2
pred	1154.18	−0.7603	0.921
nmod	2533.42	−0.9551	0.872
iobj	71.24	−1.0954	0.945
obj	1341.97	−1.1111	0.942
appos	151.25	−1.1513	0.996
nsubj	1805.06	−1.2124	0.966

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

Syntactic function	Parameter *a*	Parameter *b*	R^2
obl	2903.76	−1.2211	0.969
amod	558.25	−1.3165	0.955
advmod	161.20	−1.3288	0.952
nummod	59.27	−1.4099	0.979
det	41.37	−1.5830	0.993

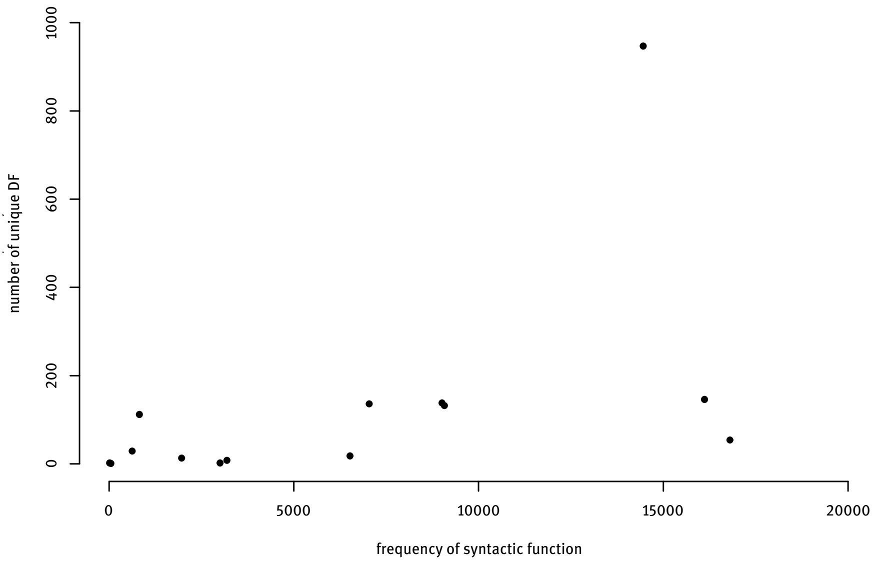

Fig. 11: Frequencies of particular syntactic functions and numbers of their unique DFs.

The hypothesis was tested using the Kendall correlation coefficient which takes the value τ = 0.505. Then the null hypothesis τ = 0 (which corresponds to no increase of the number of unique DFs when the frequency of a syntactic function increases), with the resulting p-value of 0.006. Both the relatively high correlation coefficient and the p-value do not lead to a rejection of the hypothesis. However, a deeper explanation of the result is needed. First, an extraordinary position of predicate is striking (cf. Fig. 11): its frequent occurrence is accompanied by a very high number of unique DFs which is not, however, a typi-

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

cal behaviour of other syntactic functions. From the linguistic point of view, this result indicates a high variability of syntactic contexts ruled by the predicate which is connected to its crucial role in a clause. As for the other syntactic functions, there is also positive correlation between frequency of FEW with the given syntactic function and the number of unique DFs. Specifically, if predicate is omitted, the Kendall correlation coefficient is $\tau = 0.529$ (the null hypothesis on the zero correlation is rejected also in this case), which corroborates the hypothesis, too.

5 Conclusion

The results presented in the study reveal three main findings. First, the observed regular frequency distribution of DFs in the corpus can be interpreted as a result of the least effort principle (Zipf 1949) or as an outcome of a diversification process in language (Altmann 2005). It means that this linguistic unit can be included among well-established ones and, consequently, its properties can be used in a general model of language system, such as the model within synergetic linguistic approach.

Second, there are differences among rank-frequency distributions of DFs of syntactic functions of FEWs. The differences are expressed in terms of the values of parameter *b* (see Tab. 5) - the farther from zero, the steeper the curve. Steep curves indicate that there are few dominant DFs which occur very frequently and many DFs with marginal occurrence. On the other hand, the curve for predicate does not decrease so steeply. It seems that frequencies of DFs for predicate are distributed more uniformly than for other syntactic functions which can also be seen in Fig. 11. Predicate is connected to many DFs, but the frequent ones are less dominant (with respect to their frequencies) as for other syntactic functions. The special position of predicate can be - at least partly, as predicate is mostly realized by verbs - explained by a special position of verbs, which was shown by Čech et al. (2011).

Third, there is a relation between the frequency of a syntactic function and the number of its unique DFs, see Fig. 11. For the time being, we are not able to express the relation in terms of a simple mathematical function. The reasons (one of them is a special position of predicate, but there can be many other factors at play, some of which can be stronger than the frequency) present one of challenges for future research.

If the results achieved in this study are corroborated on data from several languages, the regularities observed here can be considered language laws. In

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

such a case the laws should be incorporated into a language theory and interrelations with other language properties must be established.

It would also be interesting to investigate properties of syntactic dependencies in texts as opposed to corpora. The question whether there are typical parameter values for particular text groups (determined by genres, authors etc.) is more likely to be answered if several syntactically annotated complete texts from the same language are available.

Acknowledgement: J. Mačutek was supported by grant VEGA 2/0054/18.

References

Altmann, Gabriel. 2005. Diversification processes. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski (Eds.), *Quantitative Linguistics: An International Handbook* (pp.646–658). Berlin & New York: de Gruyter.

Bejček, Eduard, Eva Hajičová, Jan Hajič, Pavlína Jínová, Václava Kettnerová, Veronika Kolářová, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Magda Ševčíková, Jan Štěpánek & Šárka Zikánová. 2013. *Prague Dependency Treebank 3.0*. Praha: Univerzita Karlova v Praze, MFF, ÚFAL. http://ufal.mff.cuni.cz/pdt3.0/, Dec 2013.

Čech, Radek, Petr Pajas & Ján Mačutek.2010. Full valency. Verb valency without distinguishing complements and adjuncts. *Journal of Quantitative Linguistics, 17*(4), 291–302.

Čech, Radek, Ján Mačutek & Zdeněk Žabokrtský. 2011. The role of syntax in complex networks: local and global importance of verbs in a syntactic dependency network. *Physica A: Statistical Mechanics and its Applications, 390*(20), 3614–3623.

Čech, Radek, Ján Mačutek, Zdeněk Žabokrtský & Aleš Horák. 2017. Polysemy and synonymy in syntactic dependency networks. *Digital Scholarship in the Humanities, 32*(1), 36–49.

de Marneffe, Marie-Catherine, Bill MacCartney & Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC 2006*.

de Marneffe, Marie-Catherine & Christopher D. Manning. 2008. The Stanford typed dependencies representation. In *COLING: Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation*, 1–8.

de Marneffe, Marie-Catherine, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre & Christopher D. Manning. 2014. Universal Stanford Dependencies: A cross-linguistic typology. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (Eds.), *Proceedings of LREC 2014*. Paris: European Language Resources Association.

Hudson, Richard. 2007. *Language Networks: The New Word Grammar*. Oxford: Oxford University Press.

Köhler, Reinhard. 1986. *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik*. Bochum: Brockmeyer.

Köhler, Reinhard. 2005. Synergetic linguistics. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski (Eds.), *Quantitative Linguistics: An International Handbook* (pp.760–775). Berlin & New York: de Gruyter.

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM

Köhler, Reinhard. 2012. *Quantitative Syntax Analysis*. Berlin & Boston: de Gruyter.
Liu, Haitao. 2009. Probability distribution of dependencies based on a Chinese Dependency Treebank. *Journal of Quantitative Linguistics, 16*(3), 256–273.
Liu, Haitao, Chunshan Xu & Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21, 171–193.
Mačutek, Ján, Radek Čech & Jiří Milička. 2017. Menzerath-Altmann law in syntactic dependency structure. In Simonetta Montemagni & Joakim Nivre (Eds.), *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)* (pp.100–107). Linköping: Linköping University Electronic Press.
Mačutek, Ján & Gejza Wimmer. 2013. Evaluating goodness-of-fit of discrete distribution models in quantitative linguistics. *Journal of Quantitative Linguistics, 20* (3), 227–240.
Mel'čuk, Igor A. 1988. *Dependency Syntax: Theory and Practice*. Albany, NY: State University of New York Press.
Nivre, Joachim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty & Daniel Zeman. 2016. Universal Dependencies v1: A Multilingual Treebank Collection. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (Eds.), *Proceedings of LREC 2016*, 1689--1666. Paris: European Language Resources Association.
Wimmer, Gejza & Gabriel Altmann. 2005. Unified Theory of some linguistics laws. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski (Eds.), *Quantitative Linguistics: An International Handbook* (pp.791—807). Berlin & New York: de Gruyter.
Zeman, Daniel. 2015. Slavic Languages in Universal Dependencies. In Katarína Gajdošová & Adriána Žáková (Eds.), *Natural Language Processing, Corpus Linguistics, Lexicography* (pp.151–163). Lüdenscheid: RAM-Verlag.
Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison Wesley.

Brought to you by | Cambridge University Library
Authenticated
Download Date | 10/9/18 1:01 PM