Physica A 390 (2011) 3614–3623

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

The role of syntax in complex networks: Local and global importance of verbs in a syntactic dependency network

Radek Čech [a,*], Ján Mačutek [b], Zdeněk Žabokrtský [c]

[a] *Department of Czech Language, University of Ostrava, Reální 5, Ostrava, 701 03, Czech Republic*

[b] *Institute for Slavic Studies, University of Graz, Merangasse 70, Graz, 8010, Austria*

[c] *Institute of Formal and Applied Linguistics, Charles University in Prague, Malostranské náměstí 25, Prague, 11800, Czech Republic*

ARTICLE INFO

Article history:
Received 11 November 2010
Received in revised form 20 May 2011
Available online 30 May 2011

Keywords:
Syntactic network
Complex network
Verb

ABSTRACT

Syntax of natural language has been the focus of linguistics for decades. The complex network theory, being one of new research tools, opens new perspectives on syntax properties of the language. Despite numerous partial achievements, some fundamental problems remain unsolved. Specifically, although statistical properties typical for complex networks can be observed in all syntactic networks, the impact of syntax itself on these properties is still unclear. The aim of the present study is to shed more light on the role of syntax in the syntactic network structure. In particular, we concentrate on the impact of the syntactic function of a verb in the sentence on the complex network structure. Verbs play the decisive role in the sentence structure ("local" importance). From this fact we hypothesize the importance of verbs in the complex network ("global" importance). The importance of verb in the complex network is assessed by the number of links which are directed from the node representing verb to other nodes in the network. Six languages (Catalan, Czech, Dutch, Hungarian, Italian, Portuguese) were used for testing the hypothesis.

1. Introduction

The observation of syntactic characteristics of natural language by methods developed within the complex network theory [1,2] has opened up a promising way of tackling some traditional problems concerning syntax; like, e.g. the origin of syntax [3], the relation between syntax and communication needs [4,5], language acquisition [6], syntactic language universals [7,8], the origin of projectivity, i.e. the exceptionality of syntactic dependency crossing [9,10], and language typology [11–13]. Despite these achievements, there are some fundamental obscurities which (at least partly) undermine the use of complex network theory for syntactic studies. Specifically, although all observed syntactic networks based on dependency grammar formalism [14] express statistical properties typical of complex networks, the impact of syntax itself on these properties is still unclear, as has been demonstrated experimentally in Refs. [15,16]. The main problem, in general terms, is that there is an absence of any linguistic explanation of language-based networks. Up until now, the majority of language network analyses have been merely descriptive [8] and focused on global network characteristics.

The present study aims to explain (at least partly) the role of syntax in syntactic network structure. The idea is as follows: since it is well known that the shape of a network (its topological properties) is closely related to its functionality [17], we focus on the syntactic functions of network elements within a sentence (we have termed these "local" properties) and we predict the properties of these elements in the network ("global" properties). Specifically, we center on the syntactic role of a verb in the sentence and deduce how the functioning of a verb should affect both the role of verbs in a network

* Corresponding author. Tel.: +420 597091863; fax: +420 596113009.
E-mail addresses: radek.cech@osu.cz (R. Čech), jmacutek@yahoo.com (J. Mačutek), zabokrtsky@ufal.mff.cuni.cz (Z. Žabokrtský).

0378-4371/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.05.027

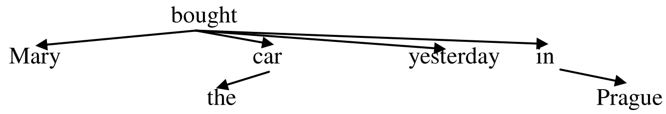

Fig. 1. The tree graph expressing the structure of sentence (2) based on the dependency grammar formalism. Links between words represent syntactic dependency relationships.

structure (Section 5) and the topology of the complex network. In short, the role of syntax in the complex network is derived from the function of verbs in the sentence. Deductions are formulated as empirically testable hypotheses, which are tested by common statistical methods (Section 8). Since it is assumed that the relationship between local and global properties expresses a universal property of language, six languages (Catalan, Czech, Dutch, Hungarian, Italian, Portuguese) are used for hypotheses testing, thus reducing the risk of language specific results.

2. The character of a verb meaning and its impact on a sentence structure

The term 'verb' is assigned to words which designate events (or states). The semantics of a verb reflects the complex character of an event (or state) in the sense that knowledge of a particular verb meaning (e.g. *buy*) involves a knowledge of the "structure" of the situation which is designated. Specifically, the verb *buy* designates situations which comprise at least two entities—a buyer and a thing which is bought. These entities constitute a *frame* in which the verb is used. Knowledge of the verb meaning thus involves knowledge of the frame. Consequently, the verb meaning (including its frame information) has decisive importance for a sentence structure—it determines the number, meaning, and grammatical form of the entities constituting the frame [18]. In the case of the verb *buy*, all grammatical well-formed English sentences containing the predicate *buy* should include a nominative subject and an accusative object, as is illustrated in sentence (1).

$$\underset{\text{[nominative, subject]}}{\textit{Jane}} \quad \textit{buys} \quad \underset{\text{[accusative, object]}}{\textit{bananas}} \tag{1}$$

In linguistics, the property of meaning that carries information on other sentence constituents is called *valency*, which is defined as "the capacity of a verb (...) for combining with particular patterns of other sentence constituents, in a similar way to that in which the valency of a chemical element is its capacity for combining with a fixed number of atoms of another element" [19, p. 4878].

The importance of verbs in grammar is also manifested in the process of language acquisition. As Ref. [20, p. 7] emphasizes, "the acquisition of verbs as single-word lexical items during the 2nd year of life is the major turning point in children's transition to adultlike grammatical competence. The grammatical valences contained in children's first verbs simply "beg" to be completed into sentences".

In sum, verbs play a crucial role in a sentence and they have a decisive impact on a sentence structure, as is illustrated in the next section.

3. Local importance of verbs

Let us term the ability of a verb to structure the sentence the *local importance* of the verb. The importance is determined both by the valency properties of the verb and by the specific role of the verb in the sentence—in a vast majority of sentences of the language a finite verb form expresses a sentence predicate (or at least a part of a predicate), which is a crucial element of the sentence meaning. The importance of the verb is expressed by the position of the verb in the graph representing the sentence structure.[1] The verb predicate is considered to be the head (governing) word of the sentence, and it is assigned as a root element of a tree graph expressing the sentence structure, as is illustrated in Fig. 1 for sentence (2).

Mary bought the car yesterday in Prague (2)

If the verb occurs in the sentence in an infinitive form (e.g. in the function of an object of the sentence), it loses only the capacity to combine with the subject; all other valency properties remain. More specifically, in sentence (3)

Mark decided to buy Mary the book (3)

the syntactic relationships between the verb *buy* and the nouns *book* and *Mary*, respectively, are ruled by the valency characteristics of the verb *buy* (see Fig. 2).

Moreover, the number of directly dependent elements of the verb is not determined only by its valency properties. The circumstances (e.g. time, place, quality) of the event are also expressed by elements which directly depend on the verb; see the adverb *yesterday* and the prepositional phrase *in Prague* in sentence (2). Ref. [22] showed that also these "free" complements, traditionally called adjuncts, are ruled by similar mechanisms as obligatory complements within the traditional valency approach [23]; the mechanism is called full valency.

[1] In this article we will follow the dependency grammar approach [14], which is one of the two most common approaches for characterizing a sentence structure in linguistics (the other one is the immediate constituency approach); for more details, see Ref. [21, 117–181], pages.

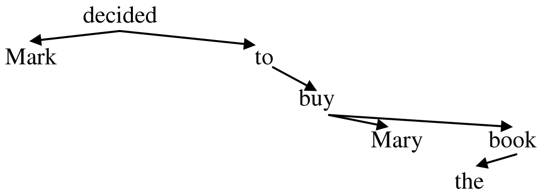

Fig. 2. The tree graph expressing the structure of sentence (3).

Therefore, since we assume that the function of elements affects the shape of the network, it seems reasonable to expect that the syntactic function of verbs should play an important role in the network expressing syntactic relationships in the language. In other words, it is predicted that verbs will occur among the most important elements of the network.

4. Definition of the global importance of a node

Let us define global importance as a property of a complex network element. A complex network is a graph which is given by a set of elements (termed *nodes* or *vertices*) and a set of connections between them (termed *links* or *edges*) [17].

The importance of a node in the network can be calculated by different ways. It is usually determined by using the following:

(a) the number of links which are connected to the node—the more links the node has, the more important it is; this number constitutes the degree of the node;

(b) the number of the shortest paths between other nodes that run through the node—this number constitutes the betweenness centrality of the node and it is usually used in social network analysis [24,25];

(c) the hub/authority weight measurement, which calculates not solely the number of links of the node but also the "quality" of connected nodes (it is determined by the number of links of connected node) [26].

Since we hypothesize that the global (topological) importance of verbs in the complex network should be caused by the valency mechanism (see Section 3), we decided to determine the importance of the node by its degree. The reason is as follows: verb valency is the mechanism which rules the behavior of verb with regard only to its *directly* dependent elements – it determines the number of elements, their meaning and form – and it has no impact on the character of other elements of the sentence (their number, meaning, and form). Consequently, it seems reasonable to consider the degree as an important indicator because it best reflects the behavior of the node of the network with regard to its *directly* connected nodes.[2]

In an oriented network, where links are directed, two kinds of degrees can be computed; the out-degree, which labels the number of links that are directed from the node, and the in-degree, which labels the number of links that are directed to the node. For an illustration, Fig. 3 shows an example of a syntactic dependency network based on Czech data containing 51 vertices; the network is oriented, which reflects the dependency formalism used for expressing the sentence structure (see Figs. 1 and 2).

The importance of the most highly connected nodes in a complex network is illustrated by their role in the structure of the network—if they are deleted, the network properties change radically and the network breaks into fragments [27].

As for the betweenness centrality, we do not assume that it should reflect directly the properties of node in the network caused by valency. The betweenness centrality takes into account not only the relationships between directly connected nodes but also the relationships among other nodes have an impact on its value. Nevertheless, we test also differences between properties of verbs with respect to degrees and betweenness centrality. The results of exploring the relationship (or better, the absence of the relationship) between local properties of verbs, as defined in Section 3, and global importance of verbs determined by measurement of betweenness centrality are presented in Section 8.

The hub/authority weight measurement is a more complex approach to evaluation of node importance. We have not found the way how to interpret it linguistically, therefore, we have decided not to apply it in the present study.

5. Global importance of verbs

The crucial question is obvious: why should verbs be ranked among the important elements of the network? The answer is less straightforward than in the case of local importance; previous studies [6,7,15,16,28] reveal that in syntactic networks, nodes with high connectivity represent the so-called function (grammatical) words, which belong to the group of closed-class words (such as prepositions, conjunctions, pronouns) and have little lexical meaning. Of course, there is a subgroup of verbs, auxiliary verbs (e.g. *be, have*), which are ranked among function words and which were detected among the highly connected words. But as far as the whole class of verbs is concerned, there is no evidence that they are more important elements of the network than other parts of speech.

[2] Of course, some valency-implied characteristics of a directly dependent element can be propagated to its children (especially by agreement rules), but we do not consider it as a direct effect of the valency mechanism.

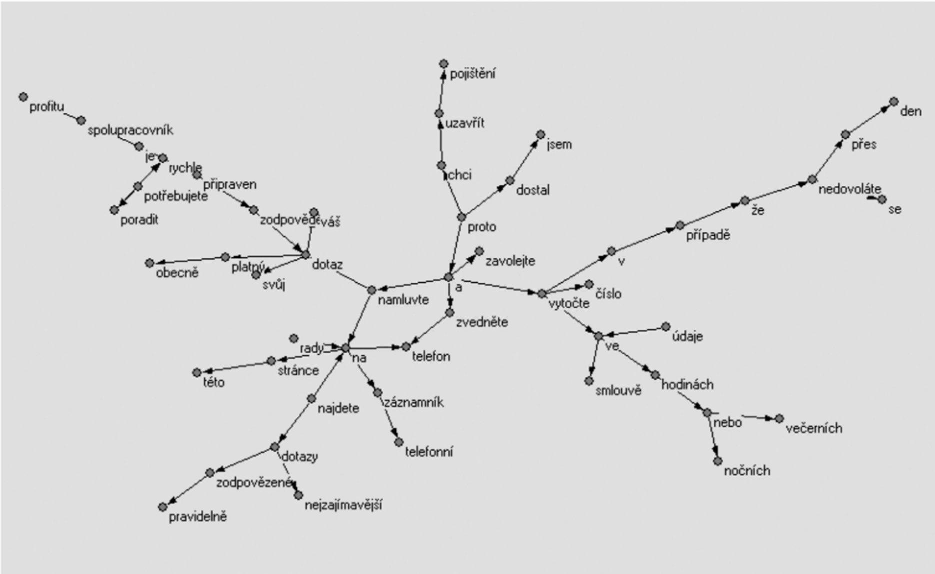

Fig. 3. An example of a syntactic dependency network containing 51 vertices.

Table 1
The distribution of parts of speech in Czech National Corpus SYN2010.

Part of speech	Token frequency	Percentage (%)
Noun	29 808 857	29.45
Verb	17 828 515	17.61
Pronoun	11 269 061	11.13
Adjective	11 083 367	10.95
Preposition	10 692 380	10.56
Conjunction	7 813 042	7.72
Adverb	7 109 130	7.02
Numeral	2 647 110	2.62
Particle	1 482 427	1.46
Interjection	67 815	0.07
Untagged	1 417 899	1.40

However, when the impact of local properties (see Section 3) is taken into account, the reasons for verbs' global importance emerge. First, each sentence usually has to contain at least one verb,[3] which is the head (governing) word and which represents the sentence predicate. This fact guarantees a relatively high frequency of verbs in any sample; for example, verbs are the second most frequent part of speech in Czech, representing about 17% of all words (tokens) occurring in Czech texts [29]; see Table 1.

However, the frequency itself does not cause the high connectivity of verbs in the network, as is shown in Section 8. Consequently, another reason is needed to explain the high connectivity of verbs. Since verb valency obligatorily "requires" the expression of directly dependent elements of the verb representing its frame in each sentence, it seems reasonable to expect that valency is one of the main causes of verbs' high connectivity. In sum, the importance of verbs in the network is deduced to be a result of a combination of (1) the relatively high frequency which is due to the predicative role of the verb, and primarily, (2) the valency properties of verbs. Importantly, each of the factors demonstrates a *syntactic* property of the language. In Section 8, the role of each property will be shown.

One can object that it is not surprising that verbs will be ranked among the group of vertices with the highest degree because the network is induced from dependency trees where a verb is a syntactic head of a graph representing a sentence. In other words, high degrees of verbs seem to be an obvious consequence of the procedure of network induction. However, the relationship between the head position in the dependency tree and a high connectivity is not straightforward—it must be emphasized that the position of head does not mean that the head automatically obtains the highest number of links in comparison to the other elements of the sentence. For example, in the sentence (4)

My nice rich friend from Prague sleeps (4)

[3] There are some exceptions, e.g., Latin or Russian, in which some sentences do not contain a verb.

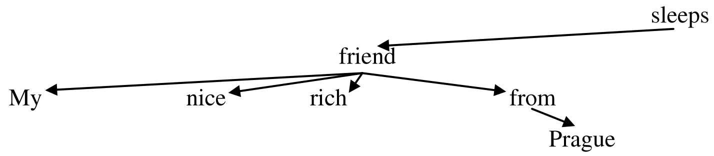

Fig. 4. The tree graph expressing the structure of sentence (4).

the verb *sleeps* has one dependent element (i.e. its out-degree equals one) while the noun *friend* has four dependent elements (i.e. its out-degree equals four); see Fig. 4.

Just for the sake of illustration, we focused on all sentences containing both verbs and nouns in the Prague Dependency Treebank [30] and observed their out-degrees. From the totality of 54,022 sentences containing both verbs and nouns, nouns have higher out-degree than verbs in 4261 cases. In other words, in almost every twelfth sentence the head position does not automatically "provide" the highest out-degree of verb in the sentence.

6. Language material and the methodology of network construction

In this section, we describe the data which we have used for building complex networks for several languages. The main source of the data was the collection of dependency treebanks originally gathered for the purposes of CoNLL 2006 and 2007 shared tasks on dependency parsing; see Refs. [31,32] for more information about the original constituency treebanks and their conversion to CoNLL dependency structures. We used data subsets for the following languages:

Catalan—data originating from the CESS–ECE corpus [33],
Czech—data originating from the Prague Dependency Treebank 2.0 [30],
Dutch—data originating from the Alpino treebank [34],
Hungarian—data originating from the Szeged treebank [35],
Italian—data originating from the Italian Syntactic–Semantic Treebank [36],
Portuguese—data originating from Floresta synta(c)tica [37].

In constructing the networks, the method developed by Refs. [28,38] is followed. The syntactic dependency network thus contains all lemmas in the corpus—the term lemma refers to a canonical word form, e.g. the infinitive form of a verb and the nominative form of a substantive, adjective, pronoun and so on (for example, the lemma of words *write, wrote, written, writes* is WRITE). Two lemmas are linked if there is a dependency relationship between them in the corpus. The links are directed in accordance with the adopted formalism, i.e. going from the head to modifier (see Figs. 1, 2 and 4). All connections between particular lemmas are counted, which means that the graph reflects the frequency of connections (and is known as a multigraph). A global syntactic dependency network is constructed by accumulating sentence structures, and the network should be viewed as an emergent property of sentence structures [7,28]. The free software *Pajek* 2.00 [39] was used for network creation and computing.[4]

7. Methodology for measuring the global importance of verbs

The global importance of verbs in the analyzed networks was assessed using the following method.

First, the ranked distribution of out-degrees is constructed, i.e. out-degrees are ordered from the highest to the lowest (the highest out-degree has the rank 1, the second highest the rank 2, etc.). Only out-degrees are counted because the method adopts a syntactic formalism which assumes that the direction is from the head to the modifier (see Figs. 1–3), and consequently this direction expresses the character of valency properties. The information on parts of speech (for the purposes of our study it is sufficient to distinguish verbs V and other parts of speech X) is kept with each out-degree. As an example we present a part of the ranked distribution of out-degrees obtained from the Catalan corpus (Table 2).

Then, the data (i.e., the value 1 occurring 26 959 times, the value 2 occurring 12 428 times, etc.) were organized into a histogram in two steps.

1. In the first step, the histogram bin width was determined according to Ref. [40] (Scott's algorithm for the number of bins in a histogram and their width is implemented in the statistical software R).
2. In the second step, the bins are adjusted if the highest frequency in the successive group created in the first step would be the same as the lowest frequency in the previous group; or, in other words, lemmas with the same out-degrees are kept in one histogram bin even at the cost of violating equal bin widths. This second step is necessary in order to avoid splitting lemmas with the same out-degree into different histogram bins — here we do not consider any other hierarchy

[4] Ref. [38] describes the process of network creation by Pajek in detail.

Table 2
Ranked distribution of out-degrees in Catalan corpus. The symbol *V* is assigned to verbs, the symbol *X* is assigned to other parts of speech.

Rank	Out-degree	Part of speech
1	26 959	*X*
2	12 428	*V*
3	8 657	*X*
4	7 445	*X*
5	5 778	*X*
...	...	...
26 217	0	*X*
26 218	0	*X*

Table 3
Proportions of verbs in histogram bins of the ranked distribution of out-degrees in corpora.

Group	CA	CZ	DU	HU	IT	PO
1	0.551	0.696	0.458	0.545	0.540	0.470
2	0.521	0.753	0.378	0.427	0.504	0.439
3	0.432	0.658	0.279	0.437	0.372	0.353
4	0.375	0.611	0.252	0.405	0.393	0.327
5	0.398	0.602	0.309	0.435	0.400	0.309
6	0.306	0.560	0.247	0.336	0.311	0.269
7	0.331	0.556	0.212	0.348	0.384	0.303
8	0.341	0.497	0.302	0.375	0.221	0.247
9	0.301	0.478	0.280	0.346	0.256	0.269
10	0.355	0.551	0.259	0.342	0.325	0.235
11	0.304	0.413	0.238	0.372	0.278	0.292
12	0.266	0.474	0.260	0.351	0.288	0.255
13	0.239	0.511	0.290	0.281	0.220	0.220
14	0.291	0.498	0.263	0.253	0.179	0.185
15	0.290	0.493	0.295	0.109	0.131	0.214
16	0.240	0.450	0.278	0.034	0.091	0.191
17	0.277	0.352	0.243	0.006	0.019	0.224
18	0.202	0.454	0.159			0.160
19	0.180	0.429	0.128			0.146
20	0.254	0.461	0.060			0.143
21	0.225	0.318				0.087
22	0.206	0.320				0.058
23	0.242	0.354				0.023
24	0.157	0.204				
25	0.195	0.179				
26	0.203	0.259				
27	0.207	0.220				
28	0.190	0.250				
29	0.156	0.238				
30	0.159	0.167				
31	0.132	0.054				
32	0.086	0.022				
33	0.042	0.014				
34	0.021					
35	0.003					

among the lemmas except for their out-degrees, thus lacking a tool to differentiate among them objectively (they are ordered alphabetically in the ranked distribution of out-degrees). We note that, especially lower histogram bins tend to widen significantly after the application of the second step.

Finally, in each group the proportion of verbs (i.e. the number of verb lemmas divided by the number of all lemmas in the respective group) is determined. The same procedure is performed also for rank–frequency distributions of word lemma occurrences in the analyzed corpora, and the obtained proportions of verbs are compared (see Section 8).

8. Results

The proportions of verbs in histogram bins of the ranked distributions of out-degrees tend to decrease. They are presented in Table 3 (throughout the paper, we use the following abbreviations for the languages: CA—Catalan, CZ—Czech, DU—Dutch, HU—Hungarian, IT—Italian, PO—Portuguese).

The proportions of verbs in histogram bins of the rank–frequency distribution of lemmas (Table 4) are more or less constant and clearly tend to attain lower values than proportions from Table 3; see Fig. 5.

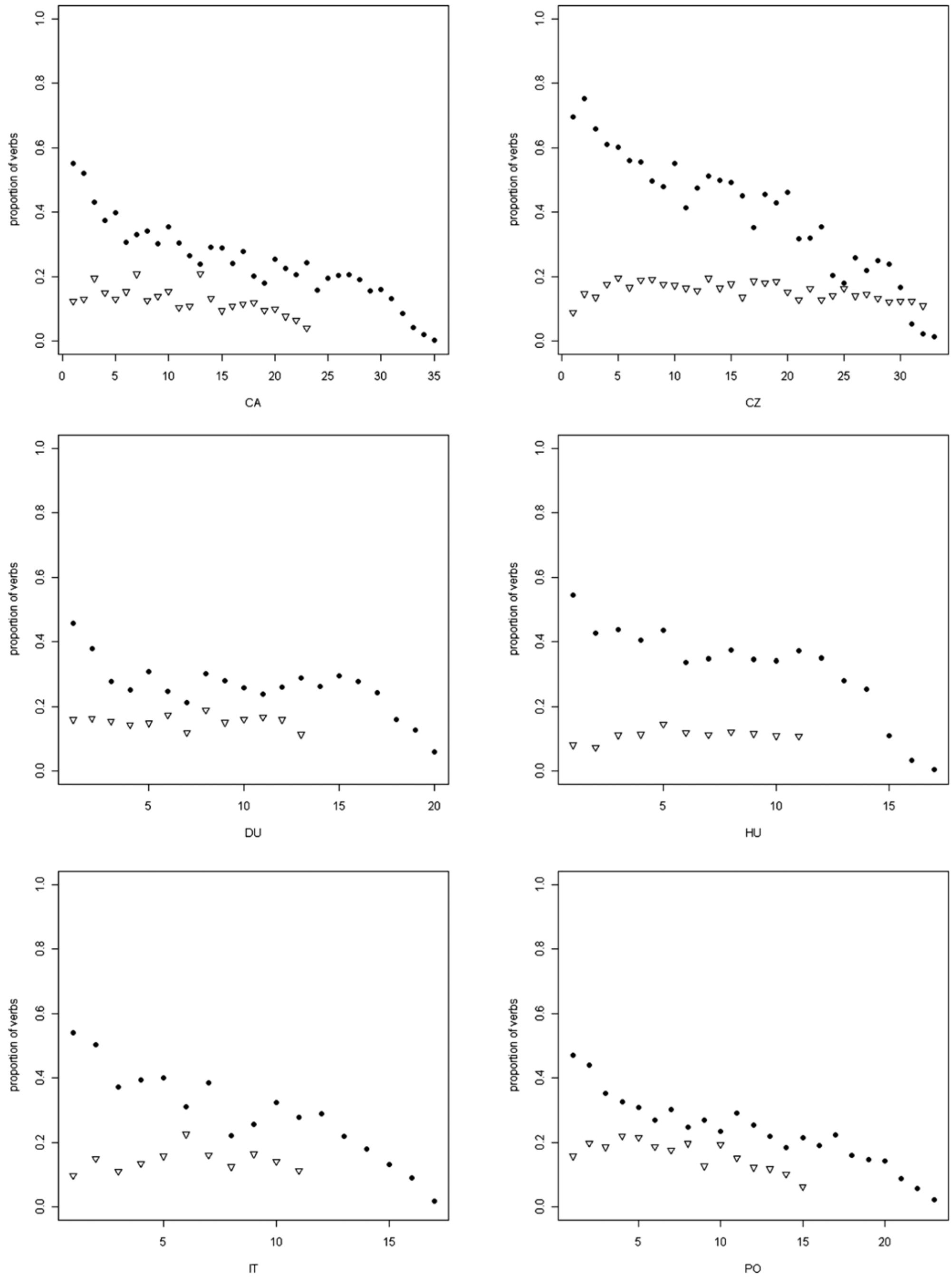

Fig. 5. Proportions of verbs in histogram bins of the ranked distribution of out-degrees (filled circles) and in histogram bins of the lemma rank–frequency distributions (empty triangles).

Differences among respective proportions from Tables 3 and 4 were tested by the two-sample t-test, wherever possible; if the data were not normally distributed, the nonparametric two-sample Wilcoxon test was applied. The p-value was lesser than 0.01 in all six cases.

An analogous comparison of the degrees and betweenness centrality[5] is presented in Fig. 6.

[5] Applying Scott's algorithm [40] to the betweenness centrality data leads – due to the low variance – to an extremely high number of histogram bins if compared with the degree distribution. Therefore, for the sake of simplicity we divide the data in the first step (cf. Section 7) into the same number of bins as the respective out-degree data, and then the second step follows.

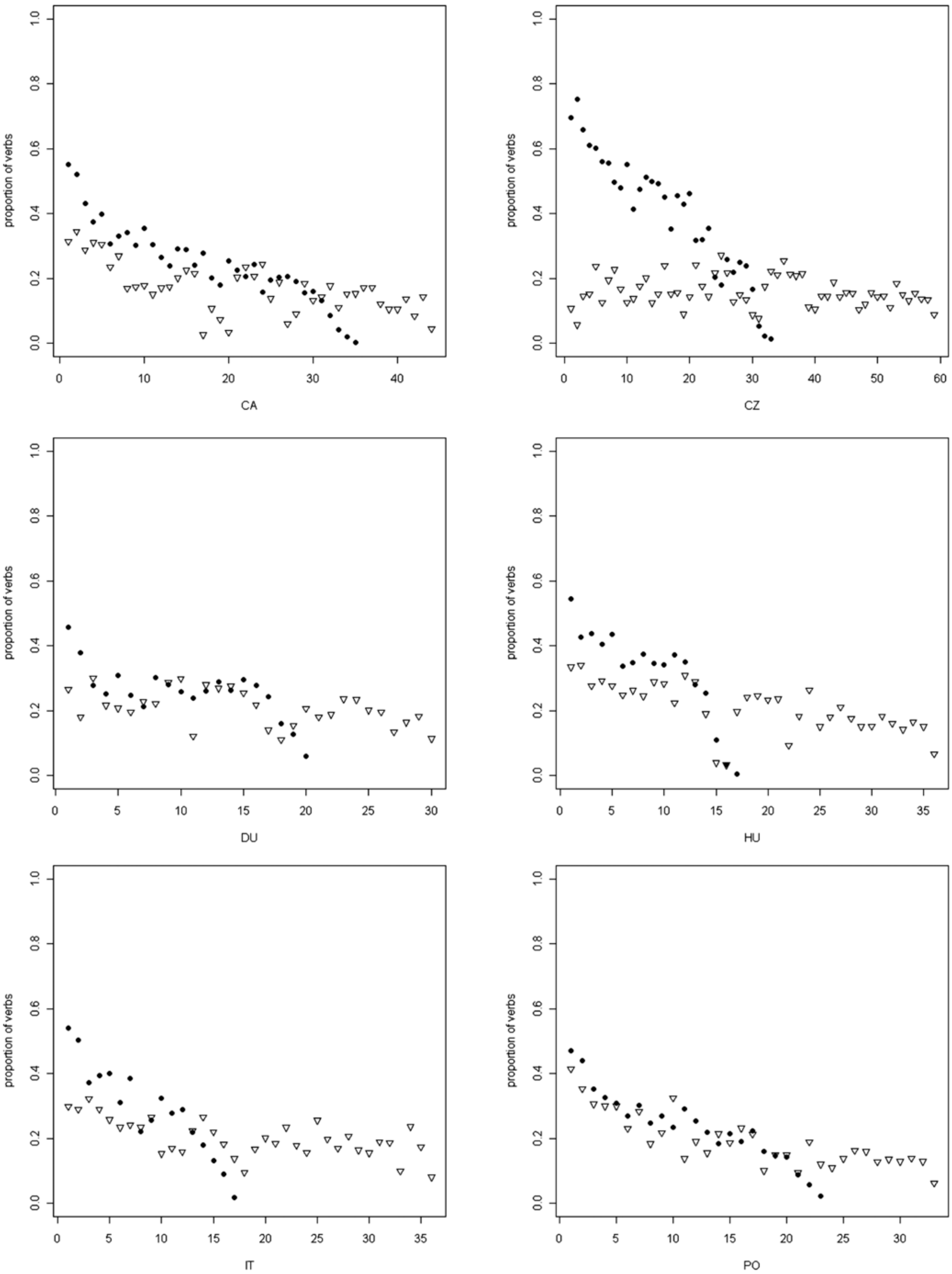

Fig. 6. Proportions of verbs in histogram bins of the ranked distribution of out-degrees (filled circles) and in histogram bins of the betweenness centrality distributions (empty triangles).

Also in this case, the differences were tested (either by the *t*-test or by the Wilcoxon test). The results are presented in Table 5.

In five out of six languages, the proportions of verbs in histogram bins of the ranked distribution of out-degrees are significantly higher than those in histogram bins of the betweenness centrality distributions. This result confirms our assumption which concerns the absence of direct relationship between valency and betweenness centrality.

In sum, the tests corroborate our hypothesis regarding the relationship between local and global properties of verbs. Moreover, the results reveal no direct impact of the frequency of verb on their position among other parts of speech with regard to the ranked distribution of out-degrees, as is obvious from Fig. 5. It indicates that valency (or full valency) should probably be considered to be the main mechanism affecting global properties of verbs in the syntactic dependency networks.

Table 4
Proportions of verbs in histogram bins of the lemma rank–frequency distributions in corpora.

	CA	CZ	DU	HU	IT	PO
1	0.124	0.089	0.160	0.081	0.097	0.158
2	0.130	0.147	0.163	0.074	0.150	0.198
3	0.195	0.136	0.154	0.112	0.111	0.186
4	0.150	0.176	0.143	0.114	0.135	0.220
5	0.131	0.196	0.149	0.145	0.158	0.216
6	0.153	0.166	0.174	0.120	0.225	0.187
7	0.208	0.189	0.120	0.113	0.161	0.176
8	0.126	0.191	0.190	0.121	0.125	0.197
9	0.139	0.176	0.151	0.116	0.164	0.127
10	0.155	0.173	0.161	0.110	0.141	0.194
11	0.104	0.164	0.168	0.109	0.113	0.152
12	0.109	0.157	0.160			0.123
13	0.209	0.195	0.114			0.118
14	0.133	0.164				0.102
15	0.094	0.177				0.063
16	0.109	0.136				
17	0.115	0.186				
18	0.120	0.180				
19	0.095	0.185				
20	0.100	0.152				
21	0.077	0.128				
22	0.065	0.163				
23	0.041	0.128				
24		0.141				
25		0.163				
26		0.140				
27		0.145				
28		0.132				
29		0.122				
30		0.124				
31		0.124				
32		0.110				

Table 5
Testing differences between proportions of verbs in degree and betweenness centrality distributions.

Language	p-value
CA	<0.01
CZ	<0.01
DU	0.023
HU	<0.01
IT	0.023
PO	0.115

9. Conclusion

The languages under observation come from different typological families, which might imply systematic differences in valency behavior of verbs. For example, rich morphology of inflectional and agglutinative languages (e.g., Czech and Hungarian in our sample) allows dropping subject pronouns, which decreases the out-degree of verbs. On the other hand, out-degree of verbs might be higher in more isolating languages due to higher frequency of auxiliaries. Nevertheless, there is a common trend clearly visible in all six languages investigated: proportions of verbs in bins of the histogram of the ranked distributions of out-degrees decrease, which means that verbs tend to belong to the most connected nodes. These proportions are significantly higher than proportions of verbs in bins of the histogram of the word lemma rank–frequency distribution; moreover, in five out of six cases the same is true also for the comparison of degrees and betweenness centrality.

Therefore, we assume that the global importance of verbs is the result of the combination of both the relatively high frequency and valency properties of verbs (and not of the frequency of verbs alone), as was discussed in Section 5.

The study reveals that the topology of a syntactic dependency network is significantly affected by the syntax of the language, namely by the valency properties of verbs. In the light of analyses [15,16] which show fundamental obscurities of the role of the syntax in the dependency network, this finding corroborates the advisability of a complex network approach in syntax research. Moreover, from the linguistic point of view, it is important that the results of the test confirm empirically the importance of the verbs and valency properties in a rather different approach to language.

Acknowledgments

R. Čech J. Mačutek, and Z. Žabokrtský were supported by GAČR (Czech Science Foundation) No. 405/08/P157—Components of transitivity analysis of Czech sentences (emergent grammar approach), by VEGA 1/0077/09, and by MSM0021620838 and MSMT ČR LC536, respectively.

References

[1] M.E.J. Newman, The structure and function of complex networks, SIAM Review (2) (2003) 167–256.
[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: structure and dynamics, Physics Reports 424 (2006) 175–308.
[3] R.V. Solé, Syntax for free? Nature 434 (2005) 289.
[4] R. Ferrer i Cancho, O. Riordan, B. Bollobás, The consequences of Zipf's law for syntax and symbolic reference, Proceedings of the Royal Society of London, Series B 272 (2005) 561–565.
[5] R. Ferrer i Cancho, When language breaks into pieces. A conflict between communication through isolated signals and language, BioSystems 84 (2006) 242–253.
[6] B. Corominas-Murtra, S. Valverde, R.V. Solé, The ontogeny of scale-free syntax networks: phase transitions in early language acquisition, Advances in Complex Systems 12 (2009) 371–392.
[7] R. Ferrer i Cancho, The structure of syntactic dependency networks: insights from recent advances in network theory, in: G. Altmann, V. Levickij, V. Perebyinis, (Eds.), The Problems of Quantitative Linguistics, Ruta, Chernivtsi, 2005, pp. 60–75.
[8] R. Ferrer i Cancho, Network theory, in: P.C. Hogan (Ed.), The Cambridge Encyclopedia of the Language Sciences, Cambridge University Press, 2010, pp. 555–557.
[9] R. Ferrer i Cancho, Why do syntactic links not cross? Europhysics Letters 76 (2006) 1228–1235.
[10] R. Ferrer i Cancho, Some word order biases from limited brain resources. A mathematical approach, Advances in Complex Systems 11 (3) (2008) 394–414.
[11] R. Éech, J. Maéutek, Word form and lemma syntactic dependency networks in Czech: a comparative study, Glottometrics 19 (2009) 85–98.
[12] H. Liu, H. Xu, Can syntactic networks indicate morphological complexity of a language? Europhysics Letters 93 (2011) 28005.
[13] H. Liu, W.W. Li, Language clusters based on linguistic complex networks, Chinese Science Bulletin 55 (2010) 3458.
[14] I. Mel'éuk, Dependency Syntax: Theory and Practice, State University of New York Press, Albany, 1988.
[15] H. Liu, F. Hu, What role does syntax play in a language network? Europhysics Letters 83 (2008) 18002.
[16] H. Liu, Y. Zhao, W. Huang, How do local syntactic structures influence global properties in language networks? Glottometrics 20 (2010) 35–39.
[17] G. Caldarelli, Scale-Free Networks: Complex Webs in Nature and Technology, Oxford University Press, Oxford, 2007.
[18] H.J. Heringer, Basic ideas and the classical model? in: J. Jacobs, A. Stechow, W. Sternefeld, T. Vennemann (Eds.), Syntax. An International Handbook of Contemporary Research, de Gruyter, Berlin, New York, 1993, pp. 297–316.
[19] D.J. Allerton, Valency grammar, in: K. Brown (Ed.), The Encyclopedia of Language and Linguistics, Elsevier Science Ltd., 2005, pp. 4878–4886.
[20] M. Tomasello, First Verbs. A Case Study of Early Grammatical Development, Cambridge University Press, 1992.
[21] R. Hudson, Language Networks. The New Word Grammar, Oxford University Press, 2007.
[22] R. Éech, P. Pajas, J. Maéutek, Full valency. Verb valency without distinguishing complements and adjuncts, Journal of Quantitative Linguistics 17 (2010) 291–302.
[23] R. Éech, J. Maéutek, On the quantitative analysis of verb valency in Czech, in: P. Grzybek, E. Kelih, J. Maéutek (Eds.), Text and Language. Structures. Functions. Interelations. Quantitative Perspectives, Praesens, Wien, 2010, pp. 21–29.
[24] S. Wasserman, K. Faust, Social Network Analysis, Cambridge University Press, Cambridge, 1994.
[25] J. Scott, Social Network Analysis: A Handbook, Sage Publications, London, 2000.
[26] J. Kleinberg, Authoritative sources in a hyperlinked environment, in: Proc. 9th ACMSIAM Symposium on Discrete Algorithms, 1998, pp. 668–677.
[27] R. Albert, H. Jeong, A.-L. Barabási, Error and attack tolerance of complex networks, Nature 406 (2000) 378–482.
[28] R. Ferrer i Cancho, R.V. Solé, R. Köhler, Patterns in syntactic dependency networks, Physical Review E 69 (2004) 051915.
[29] Corpus SYN2010: czech national corpus—SYN2010, Institute of the Czech National Corpus, Prague, 2010. Accessible at http://www.korpus.cz.
[30] J. Hajié, J. Panevová, E. Hajiéová, P. Pajas, J. Štpánek, J. Havelka, M. Mikulová, Prague dependency treebank 2.0., Linguistic Data Consortium, Philadelphia, 2006.
[31] S. Buchholz, E. Marsi, CoNLL-X shared task on multilingual dependency parsing, in: Proceedings of the Tenth Conference on Computational Natural Language Learning, New York City, New York, June 08–09, 2006, International Conference On Computational Linguistics, Association for Computational Linguistics, Morristown, NJ, 2006, pp. 149–164.
[32] J. Nivre, J. Hall, S. Kuebler, R. McDonald, J. Nilsson, S. Riedel, D. Yuret, The CoNLL 2007 shared task on dependency parsing, in: Proceedings of the CoNLL Shared Task Session of EMNLP–CoNLL, Prague, 2007, pp. 915–932.
[33] M.A. Martí, M. Taulé, L. Marquez, M. Bertran, CESS–ECE: a multilingual and multilevel annotated corpus, 2007. Available for download from: http://www.lsi.upc.edu/~mbertran/cess-ece/publications.
[34] L. van der Beek, G. Bouma, R. Malouf, G. van Noord, The alpino dependency treebank, in: Computational Linguistics in the Netherlands CLIN 2001, Rodopi, 2001, 2002, pp. 8–22.
[35] D. Csendes, J. Csirik, T. Gyimóthy, A. Kocsor, The Szeged Treebank, Springer, 2005.
[36] S. Montemagni, F. Barsotti, M. Battista, N. Calzolari, O. Corazzari, A. Lenci, A. Zampolli, F. Fanciulli, M. Massetani, R. Raffaelli, R. Basili, M.T. Pazienza, D. Saracino, F. Zanzotto, N. Nana, F. Pianesi, R. Delmonte, Building the Italian syntactic–semantic treebank, in: A. Abeillé (Ed.), Treebanks: Building and Using Parsed Corpora, Kluwer, Dordrecht, 2003, pp. 189–210.
[37] S. Afonso, E. Bick, R. Haber, D. Santos, "Floresta sintá(c)tica": a treebank for Portuguese, in: M.G. Rodríguez, C.P.S. Araujo (Eds.), Proceedings of LREC 2002, 2002, pp. 1698–1703.
[38] H. Liu, The complexity of Chinese syntactic dependency networks, Physica A 387 (2008) 3048–3058.
[39] W. de Nooy, A. Mrvar, V. Batagelj, Exploratory Social Network Analysis with Pajek, Cambridge University Press, New York, 2005.
[40] D.W. Scott, On optimal and data-based histograms, Biometrika 66 (1979) 605–610.