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.