Language Networks:

An Insight into the Language Faculty?

Final Thesis for the Master's in Cognitive Science and Language

Author: Dana Scarinci Zabaleta

Supervisor: Antoni Badia Cardús

September 2016

FIGURES AND TABLES

1. Introduction

Applying network theory to language is a very recent approach to linguistic research; meaningfully, a book released on 2015 on the topic is titled *Towards a theoretical framework for analyzing complex linguistic networks* (Mehler, 2015). "Towards" implying —as in fact is the case— that we do not have a theoretical framework yet.

Language networks aim at capturing relations among words in a very large scale, analyzing the entire lexicon or very large corpora. If successful, the allure of the network approach to language is that it has the potential of bringing new tools to fields such as linguistics and psycholinguistics.

In this Master's thesis we will approach language networks with a critical eye from a linguist's perspective. Are language networks successful in representing key aspects of language? Are the cognitive inferences made from them sound? However, before diving into the particulars of language network research, the reader will need to bear with a somewhat hard to tread introduction (sections 2 and 3) where very basic foundations of network science are laid. Since specific terminology is abundant, all terms marked with an asterisk are defined in an alphabetically organized glossary.

In section 4 we will give a critical overview of research into semantic and syntactic networks (focusing on the latter), and in section 5 we will generate and analyze syntactic networks from English and Spanish corpora, and compare our results to those reviewed. We will end in section 6 with a reflection on modelling syntactic networks.

2. Language from the perspective of complex systems

Complexity science is a relatively recent approach to scientific research that, using tools from mathematics and physics, endeavors to explain the organization and structure of systems through the study of the relations and interactions among their constitutive elements. A system, therefore, is defined as a set of elements and the set of rules or principles that guide their interaction; in a complex system the relation between primitive elements and general behavior is not straightforward; non-linear functions, stochastic methods and computational modelling are needed to grasp how the general properties of the system emerge from the interactions among the system's units themselves (and, sometimes, between them and the environment). Érdi (2008a) outlines the differences between simple and complex systems as follows:

Simple systems	Complex systems
- Single cause and single effect - A small change in the cause implies a small change in the effects - Predictability	- Circular causality, feedback loops. - Small change in the cause implies dramatic effects - Emergence and unpredictability

Table 1 Simple vs. Complex Systems (adapted from Érdi (2008a)).

Circular causality is the property whereby, in a cause-effect chain, the consequences affect the causes —A causes B, B causes C, C affects A's behavior somehow. Feedback refers to the fact that part of the output signal of a system is *fed back* to the input, influencing its dynamic behavior. Two examples of emergence are: a) consciousness, as it is a property of some cognitive systems, but neurons are not conscious themselves, and b) the performance of a team, which is not a direct consequence of one of the player's behavior (Érdi, 2008a).

The analysis of natural phenomena in terms of complex dynamical systems started in the 1970s (Érdi, 2008b) and has been steadily gaining popularity ever since; currently, it is applied to a wide variety of issues in fields ranging from physics and biology to economical and social analysis. Language has been identified as a complex system (Ellis & Larsen-Freeman, 2006; Elman, 1995; Larsen-Freeman, 1997) and studied as such from different perspectives; on the one hand, language users are defined as primitive elements and computational modelling and psychological experimental research is carried out to understand what types of interaction boost the emergence of certain language properties, such as systematicity through morpheme use (Kirby, Cornish, & Smith, 2008; Steels, 2000); on the other hand, words and other linguistic units are set as primitive elements, and different relations among them —phonological, semantic, syntactic— are expressed through networks (defined in section 3.1). The properties of the resulting network are studied and compared to other networks, in order to find underlying principles that might guide their formation and give us clues about the psychological reality of language.

In this paper I will focus on the latter, with especial interest in syntactic networks.

3. Networks and Language Networks

3.1. Networks: formal definition

A network —or a graph[1]— is a mathematical object composed of individual units — vertices or nodes (v)— and the edges (e) that link them. In Table 2 we see an example of possible units and vertices:

Network	Vertex	Edge
World Wide Web	Web page	Hyperlink
Metabolic Network	Metabolite	Metabolic Reaction
Neural Network	Neuron	Synapse
Food Web	Species	Predation
Semantic Network	Word (type)	Semantic Relation (Synonymy, Hyperonymy…)
Syntactic Network	Word (token)	Dependency Relation

Table 2 Examples of vertices and edges in particular network; adapted from Newman (2010).

Let us call the set of all nodes in a specific network V; in a social network:

$V = \{Jessica, Peter, Kevin, Loise, \dots\}$.

[1] "Graph", the original term, is used in mathematics, where the properties of these objects are studied as an end in itself. "Network", on the other hand, is the term used in fields interested in the application of graph theory to the study of other phenomena. I will use both words interchangeably.

In a language network:

$V = \{a, the, house, poetry, house, cooked, I, he, \ldots.\}$.

$V \otimes V$ is the set of all possible pairs among the elements of V; for instance: $\{(Jessica, Jessica), (Jessica, Peter), (Jessica, Kevin), (Jessica, Louise), (Peter, Kevin), \ldots.\}$.

We will call E to the subset of $V \otimes V$ in which only the connected pairs—the pairs that hold a relationship of the type we are representing, such as friendship— are included. Below we find a formal definition of two types of graph, simple and directed, using the terms just introduced.

> A simple network G is the pair $G = (V, E)$, where V is a finite set of nodes and E is a symmetric and antireflexive relation on V. In a directed network the relation E is non-symmetric. (Newman, 2010)

The "relation on V" is $E \subseteq V \otimes V$. Symmetry is the property whereby if A holds a relation with B, then B holds the same relation with A. Therefore, in n a *simple network* edges in E are unordered pairs: $e = (Jessica, Kevin) = (Kevin, Jessica)$. Antirreflexivity means that no element can hold a relation with itself. In neither simple nor directed networks will we find edges such that $e = (Jessica, Jessica)$.

The directed networks that we will consider only differ from simple networks in the fact that edges are ordered pairs: $e = (Jessica, Kevin) \neq (Kevin, Jessica)$.

In Figure 1, we can see a visual representation of a very small and imaginary friend network. A link between two nodes reflects a "friendship" relationship; since friendship is a reciprocal (symmetric) relationship, the vertices are undirected. Figure 2 is an imaginary representation of a citation network—with each vertex being a scientific paper, and vertices linking it to the papers it cites. Since citation is not a symmetric relation, links are directed (thick stubs function as arrows).

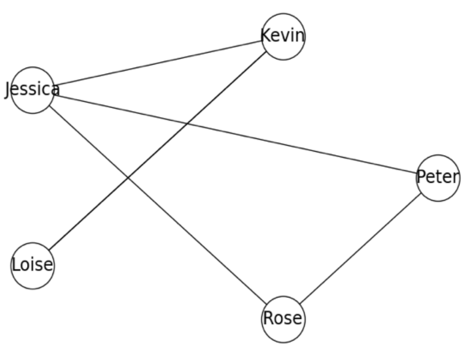

Figure 1 Example of simple network

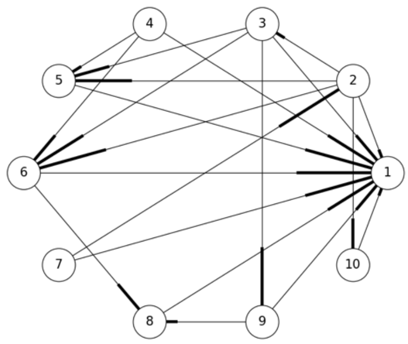

Figure 2 Example of directed network

3.2. What is it that networks capture?

The behaviour of individual units in complex systems is guided by a massively complex architecture of interacting principles; graphs —a representation of units and their relations among themselves— represent the "skeleton of complex systems" (Estrada, 2011), depicting the state of those relations at a specific point in time.

Networks capture an intermediate level, a level between the behaviour of individual units of the system —cells, atoms, people, words— and the behaviour of the whole —brain, fluids, migration, grammar. This three level distinction is known in terms of micro, meso and macro levels:

- *The micro level.* At this level the units of analysis are discrete elements. For instance, the many types of cells present in the brain or the behavior of individuals in society.
- *The meso level*. Interaction among micro units gives raise to patterns in the meso level. Continuing with the brain example, different kinds of brain tissue or functionally related neuronal networks would be appropriate units of analysis. In the social domain, behavior and interactions among social groups could be analyzed. Freeman (2000) points out that the meso is "the domain where bottom-up meets top-down," the level where we can best observe how units in the micro level give rise to macro level patterns, and how these patterns affect, in turn, the behavior of micro units.
- *The macro level*. The interaction and structure that emerges among units in the meso level is the cause of abstract system principles —macro properties— and of the emergence of aggregate behavior, such as consciousness or memory in the case of the brain, or migration movements in the case of social analysis.

Choudhury & Nukherjee (2009) explain that the micro-meso-macro distinction can be applied to the study of language. The micro units are words —tokens— used in real utterances, whereas macro units are words—types—and the grammatical generalizations that linguists study and describe. The meso level —studied through linguistic networks— captures grammar as an emergent property of token interactions and, therefore, it is expected that the features of these networks —their mathematical properties— can shed light on:

(a) Cognitive principles underlying language.

(b) The principles guiding the emergence of language as is today; if models of growth that yield networks with the same properties as language networks can be created, we will have some hints on the principles that might guide the shaping of language.

The rest of this paper will review the literature on linguistic networks, trying to assess the extent of their success. The last part of the paper will focus on syntactic networks with two questions in mind: Do syntactic networks capture relevant aspects of language? Can they shed light on the cognitive principles underlying it?

4. Overview of Linguistic Network Research

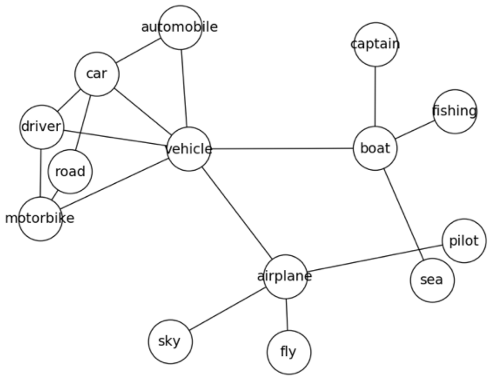

Figure 3 Piece of a semantic network

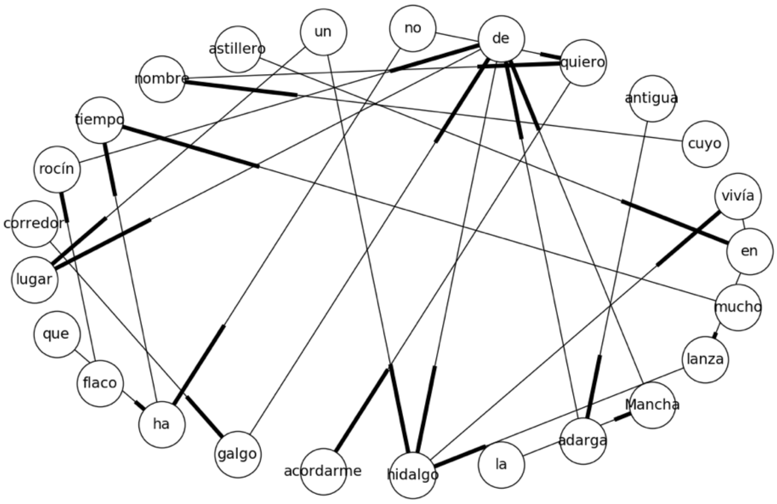

Figure 4 Syntactic dependency network of a sentence

Figures 3 and 4 are examples of the types of networks —semantic and syntactic— on which we will focus in the rest of the paper. In the semantic network, edges represent semantic proximity or relatedness, whereas in the syntactic network edges represent the dependency relations of the first sentence of Cervantes' Don Quixote. We will study the motivation and some of the mathematical properties of these objects.

A key concept will appear recurrently in this section: *small world** structure. So let us advance its meaning before we dive into the particulars of linguistic research.

Small world is a term put forward by Strogatz & Watts (1998) to name a particular type of network. After analyzing three completely different systems —the neural network of the worm *Caenorhabditis elegans*, the power grid of the western United States, and a collaboration graph of film actors— it was found that they shared two properties: just as random graphs*, these networks exhibited short *average path lengths** (see also, *path**) but, as opposed to them, they had unusually high *clustering coefficients**.

Average path length (d) measures how many steps it takes, on average, to go from any node in the network to any other node. Clustering coefficient (C) measures the probability that neighbors*—vertices sharing an edge— of a node are also neighbors of each other. (See the glossary for formal definitions of these concepts). Small world networks, therefore, are characteristic of systems where tight communities are established (reflected in the high C values), and where it is very easy to reach any element starting from any other, thanks to short *geodesic paths**.

Strikingly, this kind of organization has been found in a wide range of naturally occurring phenomena, such as food webs (Williams & Martinez, 2000), metabolic networks (Oltvai, Barabási, Jeong, Tombor, & Albert, 2000), cellular networks (Bhalla & Iyengar, 1999; Kohn, 1999), the World Wide Web (Broder et al., 2000), scientific research citation networks (Newman, 2001) and, of course, language networks. A general question for researchers in complex systems, outside the scope of this paper, is why so many self-organizing systems exhibit small word structures. Here, we will only be interested in a narrower question: what linguistic or cognitive principles explain the emergence of small networks in different representations of different language phenomena?

4.1. Semantic Networks

Semantic networks are abstract representations of one aspect of semantic knowledge: relations among words. In semantic networks vertices are words (types) or meanings, and edges represent semantic relations —such as synonymy or hypernymy (as in Figure 3 above). Whether this representation mirrors or abstracts psychological structures is a question for future research; however, it is interesting that it finds a parallel in a psycholinguistics model of how the mental lexicon is organized and accessed: the spreading activation model.

The spreading activation model has its origins in Collins & Quillian (1969) and Quillian (1967) and has a theoretical linguistics counterpart in relational theories of lexical semantics, for which the meaning of the word is defined as "the set of meaning relations in which it participates" (Geeraerts, 2010). According to this model, presented in computational terms, every lexical item in the mental lexicon —the stock of words in long-term memory from which we draw in the construction of phrases and sentences— is stored as set of pointers to words that name their synonyms, their properties or the superset they belong to, to name a few; these pointers constitute a word's meaning. Although the exact form that lexical items take in the mental lexicon is a matter of debate and diverging theories coexist (see Elman, 2004; Seidenberg & McClelland, 1989; Smith, Shoben, & Rips, 1974), the fact that words form a web of connected components is less contested.

The evidence for this fact comes mainly from semantic priming effects, the basis of Collins & Quillian's (1969) study and one of the few robust effects found in the psycholinguistics literature (for recent reviews see Hutchison (2003) and Lucas (2000)). Semantic priming can be found, for instance, in lexical decision tasks, in which participants are asked to judge whether a word is part of their language's lexicon or not. In these tasks, it is consistently observed that target items are recognized faster when a semantically related *prime* word is presented just before it. Moreover, a neurophysiological response of surprise (the N400 potential) is also weaker in this case (Bentin, McCarthy, & Wood, 1985).

Researchers agree that priming occurs because the prime partially activates related concepts and thus facilitates target words' processing. If this is indeed the case, it would seem that network modelling and psycholinguistic research could go hand in hand.

It is not surprising, therefore, that the first investigations into language networks used the WordNet Lexicon, a lexical database developed by psycholinguists that allows to explore lexically and semantically related words. Sigman & Cecchi (2002) studied the global structure of WordNet and how it changed when adding and subtracting different semantic relations.

WordNet is composed of nouns, verbs, adjectives and adverbs. Following relational theories of lexical semantics (mentioned above), the sense of each word —a string of characters— is represented as a set of synonyms (synsets); synonymy and antonymy relations are considered for all word classes, whereas others are restricted: hypernymy and meronymy apply only to nouns, and troponymy (manners of an action: *march, walk; whisper, speak*) and entailment (*marry, divorce; drive, ride*) only to verbs (Fellbaum, 1998; Miller, 1995). For the sake of simplicity, Sigman and Cecchi only analyzed the noun web: each vertex was a noun and edges represented semantic relations with other nouns. Four relationships were considered: hypernymy, antonymy, meronymy, and polysemy. Since the graph was simple all relations had to be symmetric; therefore, hypernymy also reflects hyponymy, and meronymy, holonomy.

The skeleton was always the hypernymy tree, and eight different graphs were generated by adding other relation edges[2]. Moreover, for comparison purposes, semirandom graphs were also created by randomly generating an amount of edges equal to the compared graph. All networks contained 66,025 nodes, only connections among them changed.

The resulting graphs were analyzed in terms of the following measures:

- *Characteristic length (CL).* Defined in the paper as the "median of the distribution of average minimal lengths across all vertices in the graph." The minimal length of each node is the minimum path from that node to every other node in the network and, therefore, characteristic length gives an approximation of how separate two meanings usually are —it is a measure similar to *average path length**.
- *Clustering coefficient**.
- *Traffic*, as they call it, or *betweenness centrality*.* The betweenness centrality of a node *v* is the number of shortest paths among all nodes that go through *v.*

It was found that graphs capturing hypernymy and other relations have larger *CL* (longer paths between nodes) than semirandom graphs with the same number of edges, with only one exception. When polysemy is present, semirandom and semantic graphs have similar *CL* values: this measure goes from 11.9 to 7.4, on average, when introducing polysemy links to the network. Moreover, polysemy turns a considerable amount of vertices into highly connected points, which causes clustering coefficient to increase, on average, from

[2] List of all graphs generated and analyzed, with hypernymy represented as I, antonymy as II, meronymy as III, and polysemy as IV: I, I-II, I-III, I-IV, I-II-III, I-II-IV, I-III-IV, I-II-III-IV.

0.0002 to 0.06. What this means is that it is only *polysemy* that generates small world* networks.

A difference in the *traffic* measure was also found. Without polysemy, the vertices with more traffic are those naming taxonomic groups —since finding a path between distantly related meanings, such as *dog* and *oak* involves travelling up to a common taxonomic group, e.g. *life form*. However, when polysemy is introduced, traffic shifts to "central abstract meanings" with high polysemy, such as *head* or *line.*

Polysemy completely alters the structure of semantic networks, making them easier to navigate; meanings become closer to each other thanks to the most polysemous concepts, the hubs of the network. The insight of this paper is that it suggests a very central role for polysemy in language: facilitating navigation across an immense lexical network.

The second highlight in semantic network research is Steyvers & Tenenbaum (2005), who developed a model of network growth that reproduces the structure of the three semantic networks they analyzed:

- The whole WordNet Lexicon (Fellbaum, 1998; Miller, 1995), reproduced as an undirected network.
- Nelson, McEvoy, & Schreiber's (1998)'s free word association norms. These norms were constructed by having participants respond to stimulus words with the first word that came to mind. They had around 6,000 participants and 5,000 stimulus words, and their database contains more than 750,000 words. A directed and an undirected network were constructed and analyzed.
- *Roget's Thesaurus of English Words and Phrases* (1911). This thesaurus builds a *bipartite network**, containing two types of vertices: word nodes and semantic category nodes. For a simpler analysis, the bipartite graph was converted into a simple graph by connecting words that shared at least one class.

Despite the different nature of the networks' sources, all of them turned out to be small-world* networks with the properties enumerated below. Shared characteristics across semantic networks suggest to the authors the existence of some underlying cognitive or linguistic principle guiding their organization.

- Short *average path lengths** relative to the network size.
- High *clustering coefficients**. "The associates of a word tend to also be directly associated a significant fraction (approximately 1/6) of time", indicating that related words are usually part of closely knitted communities.
- *Sparsity,* meaning that nodes, in general, are directly connected to a very small percentage of nodes in the network.
- *Connectedness.* Every network contains a single large component* that includes most nodes. That is to say, it is almost always possible to find a path* between any two words in the network.
- *Power-law* degree distribution**. The *degree** of a node (k) is its number of edges; the *degree distribution** $P(k)$ of a network is the probability that a randomly chosen node will have degree *k.* These semantic networks show degree distributions such that $P(k) \approx k^{-\gamma}$, where γ varies between 3.01 and

3.1. This means that vertices that are more connected are the scarcest, and vice versa. This kind of degree distributions are called *power law** or *scale free**.

Regarding models of network growth that could fit these characteristics, the classic one for scale free* networks is that of Barabási & Albert (1999). Their model, based on the self-organizing principle of *preferential attachment*, creates networks with an exponent of 2.9 ± 0.1, just as the semantic networks analyzed. Their network starts as a set of edges of size m_0; and at each time step, a new node is added, with m ($\leq m_0$) edges to other existing vertices.

Preferential attachment is the stochastic principle by which the existing nodes are chosen for establishing new connections. The probability of the new node connecting to existing node i is directly proportional to its *degree** k_i at a given time step.

$$\mathrm{P}(k_i) = \frac{k_i}{\sum_{j=1}^{m} k_j}$$

Unfortunately, the clustering coefficient values of networks produced by this model—around .02— are significantly lower from the C values of actual semantic networks —around .186, making it unfit to model semantic network's growth.

When trying to adapt Barabási and Albert's model to semantic networks, Steyvers and Tenenbaum rely on a particular insight from language research into child word acquisition (R. Brown, 1958; Carey, 1978, 1985): *differentiation*. It seems that one particularly productive manner in which children acquire new words is by nuancing —differentiating— already existing concepts.

Just as Barabási & Albert's (1999)'s model, Steyvers and Tenenbaum's starts with a small network of size m_0 and adds a new node connected to the existing network through m_0 edges at each time step. However, in order to implement differentiation, the edges will only be established with the neighborhood of a chosen node, selected through preferential attachment. Moreover, each node has a *utility* variable —implemented as usage frequency, following Kucera & Francis's (1968) norms— that modulates the probability of it being the target of new connections once the differentiating node has been chosen. They build two models, for directed and undirected networks, with the only difference that in the directed model, the direction of the connection is chosen randomly, and find that both of them reproduce the structure and characteristics of the networks presented above.

Despite its abstractness and limitations —such as the fact that new nodes never create connections between unconnected neighborhoods or the fact that semantic relations may change or disappear, this model may be successful in recreating an aspect of individuals' lexical acquisition (and, the authors suggest, might even also be relevant to analyze communities' lexical evolution). The model predicts that concepts acquired earlier will show the highest connectivity. If we assume that the number of connections of an item in the mental lexicon is proportional to its retrieval ease, then this model's predictions can be tested through psychological experiments measuring response times to words acquired at different stages of life.

Such experiments have been independently carried out and show the expected correlation. In picture-naming tasks, Carroll & White (1973) and Morrison, Chappell, & Ellis

(1997) found that objects that were estimated to have been learned earlier were also named faster. In lexical decision tasks —where participants have to decide whether a word belongs to their language or not— Turner, Valentine, & Ellis (1998) found that speed correlated with age of acquisition in the auditory and visual modalities. In word naming tasks —in which participants have to read the words presented, Gilhooly & Logie (1981) observed again a correlation between age of acquisition and naming latency—the time elapsed between the presentation of the word and the start of the articulation; moreover, Brown & Watson (1987), in a multiple regression study, concluded that age of acquisition was a better predictor of naming latencies than other variables such as spoken word frequency, written word frequency, and rated familiarity. In a more recent study on spoken word recognition, Garlock, Walley, & Metsala (2001) also found age-of-acquisition effects in groups of preschool and elementary-school children, as well as adults.

As we have seen, despite it being extremely young, research into semantic networks has yielded interesting insights around the mental lexicon: the importance of polysemy, on the one hand, and a plausible model of individual's lexical development, on the other. Moreover, it is crucial these advances both feed from psycholinguistics research and are able to feed back to it.

If this field is to move forward, more accurate lexical networks could be constructed and studied by adding and further categorizing relations in accordance with psycholinguistics findings. For instance, Garlock et al. (2001) also found neighborhood density effects in the three groups analyzed. Neighborhood density is defined as the number of words that differ from the target word by a phoneme addition, deletion or substitution (Metsala, 1997); this finding suggests that lexical networks could be improved by adding connections among phonologically —maybe even orthographically— similar items. Furthermore, Lucas (2000) observes different size effects according to type of relation: functional relations —usually instrument/item pairs, such as *knife/cut*— seem to be the strongest, followed by synonymy/antonymy; perceptually related pairs —objects of similar characteristics, such as *snake* and *river*, showed the weakest effects. *Weighted networks** could be constructed so that edge weights mirrored the fact that links between lexical items have different strengths in the mental lexicon.

4.2. Syntactic Networks

As opposed to the constructed nature of semantic networks, based on linguists and psychologists' efforts to organize the lexicon, syntactic networks try to capture the relations that words establish among themselves in actual language in use. Two methods have been proposed for this purpose: the co-occurrence or n-gram approach (Ramón Ferrer-i-Cancho & Solé, 2001), and the syntactic dependency approach (Ramón Ferrer-i-Cancho, Solé, & Köhler, 2004) .

The former considers only the simplest relation, contiguity or proximity. Two words —vertices in the network— are linked by an edge if they are consecutive in a text, or are separated by a distance lesser or equal than a stipulated number, usually 2 —this distance would capture the relation between *buy* and *lamp* in *buy red lamps*, but not in *buy a red lamp*. In a more restricted version of this method, links are created only if the words co-occur with higher than chance probability. Figure 5 represents a co-occurrence network of Don Quixote's first paragraph.

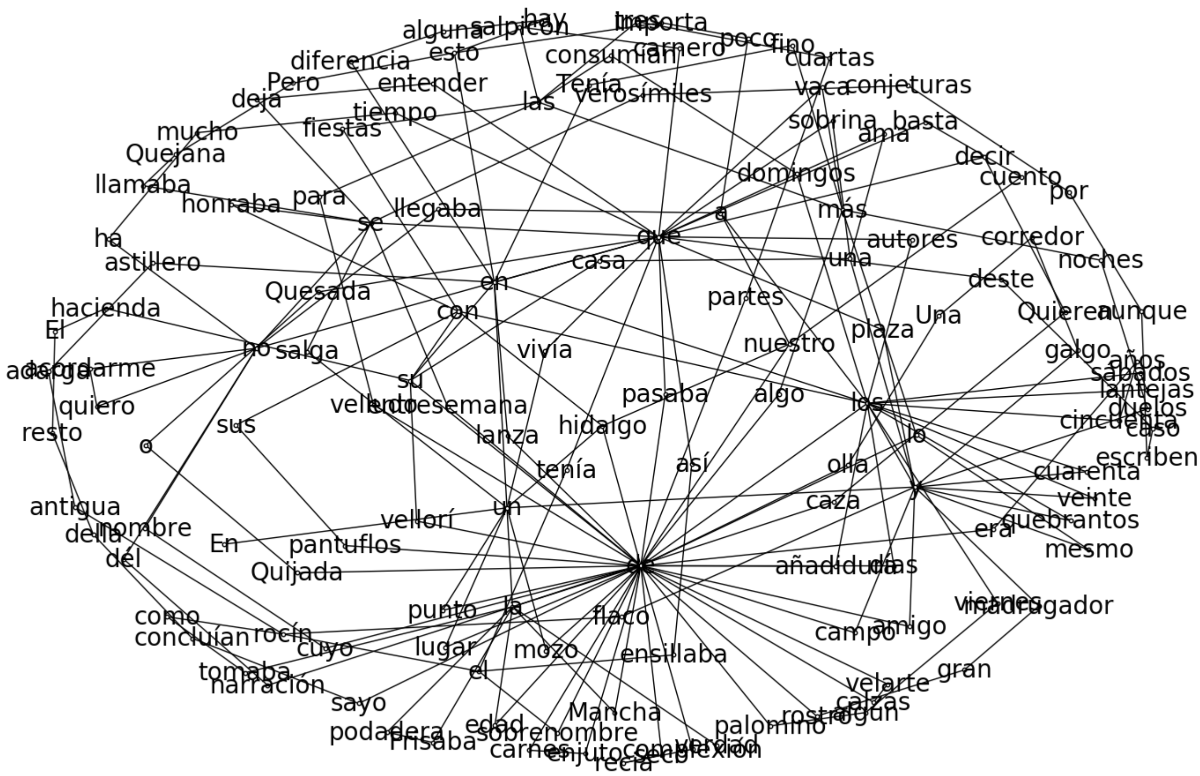

Figure 5 Co-occurrence network

Syntactic dependency networks use syntactically annotated corpora (treebanks), so that connections between words are grammatically grounded. Up to the present, only treebanks annotated according to the dependency grammar tradition have been explored.

The family of dependency grammars, whose basis was established in Tesnière (1959), functions under the premise that syntax is based on word relations; therefore, their main operator is the binary relation between word forms (word_A is linked to word_B). On the contrary, the mainstream phrase-structure grammar is focused on constituency relations, whose main operator is set inclusion (a word belongs a to phrase, which in turn belongs another phrase, etc.) (Melčuk, 1988).

Figure 6 offers a clear visual insight into the preference for one framework over the other: phrase structure grammar would entail adding hard to motivate phrase nodes to a network, whereas dependency grammars straightforwardly reflect relations between words. Moreover, since dependency structure can be interpreted in terms of phrases[3] (Hudson, 2007; Melčuk, 1988), choosing one over the other does not imply a radical difference.

[3] "each word that has at least one dependent is the head of a phrase which consists of that word plus (the phrases of) all its dependents" (Hudson, 2007)

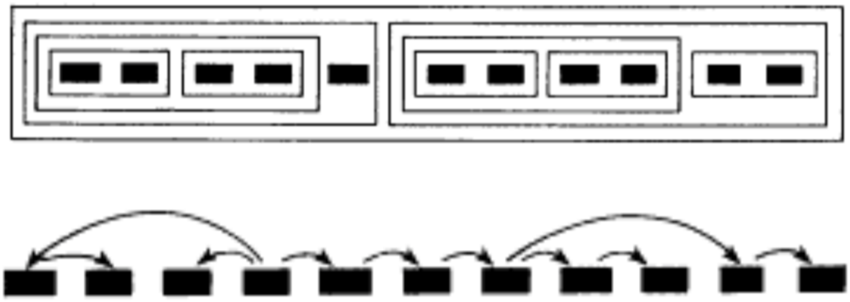

Figure 6 Phrase structure versus Dependency structure (Melčuk, 1988)

4.2.2. What do syntactic networks represent?

Syntactic networks use real texts as their source material, and represent the connections among words that are established in them. Their investigation with the purpose of learning about language as a cognitive phenomenon rests on premises that are not uncontested: either the traditional division between *langue* and *parole,* advanced by Saussure (1972) is non-existent, or both these concepts are interdependent and interrelated aspects of the same cognitive phenomenon.

The division between competence and performance is a basic pillar in the generative grammar tradition, according to which only the former can shed light into the cognitive mechanisms of language:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-communication, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of this language in actual performance. (Chomsky, 1965)

Syntactic networks are of no use to this framework. However, streams that view grammar as constantly evolving system emerging from use also exist. A first highlight in this tradition is Hopper's (1987) Emergent Grammar proposal, according to which each linguistic production contributes to an ever evolving grammatical system and, therefore, "the linguist's task is in fact to study the whole range of repetition in discourse, and in doing so seek out those regularities which promise interest as incipient sub-systems."

This line of research sees language as a complex adaptive system (Bybee, 2010) and has its most outstanding exponent in the usage-based framework, examples of which are Tomasello's (2003) account of first language acquisition as a process of reading intentions and identifying patterns through general cognitive skills such as categorization or analogy, or Construction Grammar, which views grammar as a system of interrelated constructions — form-meaning pairings— acquired through exposure to the input thanks to general cognitive skills, as well as pragmatic and processing constraints (Goldberg, 2006). Computational models of syntactic evolution that understand language as a complex system also exist (Nowak & Krakauer, 1999; Nowak, Plotkin, & Jansen, 2000).

At first sight, it seems that these approaches to language could benefit from what syntactic networks research has to offer. However, just as it was easy to accept that semantic networks represent an aspect of lexical knowledge (even if the representation may be an abstraction of the real mental lexicon), it is difficult to pinpoint what exactly syntactic networks represent. They cannot be considered a representation of the universe of constructions in the construction grammar sense[4], since not only constructions but relations among all words in a corpus are imprinted in the network.

Cong & Liu (2014) suggest that, apart from static networks such as those that semantic networks represent, dynamic networks can be constructed from instances of real language use that inform us about "the complexity of language use" and "language subsystems;" it is not clear what it is meant by that. In a reply to their paper, Hudson (2014) tries to clarify the topic as follows:

> The leading idea here is that the word-tokens of usage (actual speech events) become temporary parts of the permanent network, so [...] the dynamic and static networks are actually a single network which is constantly changing. The permanent part is what we use in both creating and interpreting word-tokens, and the dynamic part feeds back into the permanent part as we learn new patterns and absorb the statistical properties of experience (Hudson, 2014)

It is still unclear, in my opinion, what kind of cognitive representation these authors have in mind. Hudson's proposal of a temporary network that is part of a static network, although appealing, is hard to integrate with current syntactic networks, where only the temporary part is observed.

Alternatively, we can go back to Choudhury & Nukherjee's (2009) explanation of how the micro-meso-macro level distinction applies to language that was presented in section 3.2., according to which syntactic networks are simply a representation of word interactions found in instances of real language in use.

The distinction between "cognitive representation" and "representation of interactions" is not trivial, since conclusions drawn from the same network under different assumptions about its nature cannot always coincide, as we will see later. Analysis of the former informs us about the structure of some cognitive entity, whereas from the emergent properties of the latter we may be able to infer something about the cognitive principles guiding syntactic communication.

Syntactic networks will certainly reflect certain statistical properties of texts and show emergent properties of word interactions. However, without an answer to the question "What do networks represent?" it will be difficult to draw conclusions from their analysis. Let us move on to recent results from syntactic network research, but always bearing in mind that this important matter is far from settled.

---

[4] Although algorithms based on co-occurrence networks have been successful in extracting grammatical constructions. However, as opposed to the syntactic networks studied, these algorithms take into account strict sentence order, as well as the starts and ends of each sentence (Solan, Horn, Ruppin, & Edelman, 2005).

4.2.3. Analyzing syntactic networks

Ferrer-i-Cancho & Solé (2001) constructed the first simple co-occurrence network using most of the British National Corpus (2001) and found it had a small word* structure and a scale free* degree distribution*, just as semantic networks. Similar results were found for English, French, Spanish, and Japanese networks in Milo et al. (2004).

However, since these networks have the disadvantage of both capturing many spurious relations and missing many relevant syntactic relations, soon Ramón Ferrer-i-Cancho et al. (2004) resorted to treebanks annotated following dependency grammar conventions in order to build their networks.

They analyzed a Czech, a Romanian and a German corpus and found they shared the following properties:

- *Small world* structure*
- *Power law degree distribution**
- *Disassortative mixing.* Assortative mixing is the property of a network where nodes of a type tend to be connected to nodes of the same type. Disassortative mixing is the opposite property. In this paper they consider two types of nodes: highly and weakly connected, and find that the highly connected nodes tend to be connected to nodes with few connections, but not among each other.
- *Hierarchical organization*. Ravasz & Barabási (2002) had shown that networks modelling real phenomena —such as the World Wide Web, social and language networks, characterized by high clustering coefficients and power law degree distributions— are hierarchical in the sense that "small groups of nodes organize in a hierarchical manner into increasingly large groups". For an intuitive illustration, see Figure 7 and its original explanation.

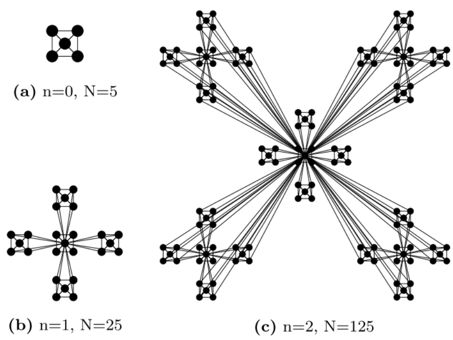

FIG. 1: The iterative construction leading to a hierarchical network. Starting from a fully connected cluster of five nodes shown in **(a)** (note that the diagonal nodes are also connected – links not visible), we create four identical replicas, connecting the peripheral nodes of each cluster to the central node of the original cluster, obtaining a network of $N = 25$ nodes **(b)**. In the next step we create four replicas of the obtained cluster, and connect the peripheral nodes again, as shown in **(c)**, to the central node of the original module, obtaining a $N = 125$ node network. This process can be continued indefinitely.

Figure 7 Evolution of a hierarchical network (Ravasz & Barabási, 2002)

Let us explore the last two properties. It was also found that the degree* of a node correlated positively with the frequency of the word it represented, and most frequent words were function words —prepositions, determiners, etcetera. It is hypothesized that disassortative mixing could be due to the fact that function words are not usually linked to each other: links tend to go from prepositions to verbs and nouns, and edges from determiners always point to nouns.

Hierarchical organization is observed when the nodes with higher degree show low clustering coefficients—the neighborhoods they belong are not very tight, since they are so big. Ferrer-i-Cancho et al. (2004) hypothesize that hierarchical organization of the network might reflect the hierarchical structure of syntax. However, this claim is left entirely unexplained. Let us reflect about this assertion. If this was indeed the case, the nodes with higher k should be those of heads in the dependency representation, that is, verbs and nouns, in that order (but only those with high frequency, since low frequency items cannot establish as many connections). We will have a chance to check this fact in section 5.6, but for now we can note that the authors reported that words with higher degree were function words, which usually occupy very low positions in the hierarchical organization of sentences.

What other conclusions could we draw from the reported properties of syntactic networks? Do they have implications about the cognitive basis of language or about language evolution?

It is unclear, and authors' suggestions are always tentative. As with semantic networks, authors point out that small world structure implies easier navigation across the web of connected nodes and Ferrer-i-Cancho et al. (2004) even put forward a function for closed category words: facilitating navigation, just as was proposed for polysemic words in semantic networks.

First, this conclusion can only be drawn if we consider syntactic networks to be cognitive representations of an aspect of language knowledge. If they are merely representation of interactions—even if we consider that these interactions have causal links to cognitive representations— proposals about navigation are hard to assume. Second, imagine a syntactic network generated from all the input and output a person has been exposed to throughout their lives: articles and prepositions are likely to be connected to almost their entire lexicon; how useful for navigation can a node be if it is connected to virtually *everything* in the network?

On a different topic, Hudson (2014) proposes that the scale-free property in these networks derives from the fact that in learning "the rich get richer," that is, the most frequently used words and structures become more strongly connected as we "try to fit new experiences into our existing knowledge." This reminds us of growth through preferential attachment presented in the semantic network section, and could fit both interpretations of what syntactic networks represent.

As we have seen, neither the foundations of nor the conclusions from syntactic network research are as strong as those its semantic counterpart. However, the fact that networks of corpora of different sizes, from different languages and annotated according to different paradigms all share many properties cannot be spurious; the current uncertainties might only be due to the fact that this is a very young field of study.

4.2.4. Setbacks

Unfortunately, the most highlighted properties of syntactic networks— small-worldliness and power law degree distributions— have been observed in networks generated from linguistic material in which different forms of randomness were introduced. For instance, Liu & Hu (2008) created three networks using a Chinese dependency-annotated treebank (no reference for the corpora is given):

1. A regular syntactic-dependency network.
2. A network generated by randomizing the links within each sentence of the corpus, but respecting two conditions: that there is a single governor for each word and a single root —head— for each sentence.
3. A network created by randomizing the links within each sentence respecting four conditions, the two previously mentioned and (a) connectedness: there are directed paths from the vertex labeled root to all other vertices, and (b) continuity: no crossing dependencies are allowed. This condition was added because it has been observed that languages display very few crossing dependencies (Ferrer-i-Cancho, 2006).

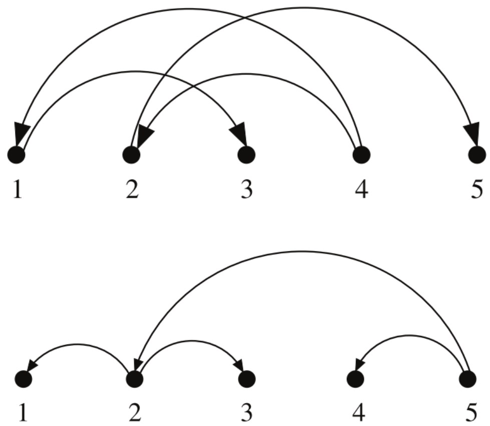

Figure 8 Crossing versus non-crossing dependencies (Liu & Hu, 2008)

They found that both true syntactic networks and randomized syntactic networks are scale-free small-world networks. Of course, one might argue that these allegedly random networks were not truly random, since conditions imposed on their construction are precisely the syntactic laws observed natural language.

However, a second study (Krishna, Hassan, Liu, & Radev, 2011) comparing English, French, Spanish and Chinese showed similar results despite analyzing truly randomized texts. They build 8 bigram networks, 4 of them lemmatized and 4 non-lemmatized. For each of the groups, they generated these four networks:

- Bigrams extracted from the original texts in two ways:
  - all bigrams
  - only highly associated bigrams
- Bigrams extracted from texts randomized in two ways:
  - words are permuted randomly in each sentence
  - words are permuted randomly in each document

All of the 8 networks show scale-free degree distributions and small-world properties. These results lead the authors of both articles to conclude that small-worldliness and power-law degree distribution are necessary but not sufficient conditions for syntactic networks, and that these properties may be related to other properties of language, possibly the frequency distribution of words stated by Zipf's law.

Zipf 's law states that if we create a list ranked by frequency of all the words in a corpus, the frequency $P(i)$ of a given type —where i is its rank— will approximately correspond to: $P(i) = p_1 i^{-\alpha}$, where p_1 is the frequency of the highest ranking word and $\alpha \approx 1$ Ferrer-i-Cancho & Solé (2001). In short, the probability of a word is inversely proportional to its rank, meaning that the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etcetera.

To these claims, Ferrer-i-Cancho (2014) responds that "from a non-reductionist view taking into account grammaticalization processes, word frequency and syntax are hard to separate;" he is stressing one of the main tenets of usage-based theories: use as the origin of syntactic patterns. Even if small-worldliness and scale-free degree distributions are a consequence of word frequencies, word frequencies could be a consequence of how patterns —syntactic or otherwise— take root through cognitive processes.

Despite the somewhat disappointing results obtained from randomized texts, difference in the global properties of networks do appear. Both articles find that the *clustering coefficient** (C) —the probability that two vertices that are neighbors of a given vertex are also neighbors of each other— seems to increase with the networks increasing levels of randomness. The values of C in Liu & Hu (2008) are shown in Table 3, and in Table 4 for some of the languages analyzed in Krishna et al. (2011).

Regular dependences (1)	Most restricted random dependencies (3)	Least restricted random network (2)
0.128	0.175	0.185

Table 3 Liu & Hu's (2008) randomized dependencies C values

	Non-randomized	Randomized sentences	Randomized texts
English	0.5520	0.5910	0.6345
English stemmed	0.5991	0.6582	0.7143
French	0.4550	0.5467	0.5612
French stemmed	0.4758	0.5745	0.6172
Spanish	0.4017	0.5315	0.5094
Spanish stemmed	0.5044	0.5812	0.5844

Table 4 Krishna et al.'s (2011) randomized dependencies C values

We know that a defining property of small-world networks is that they have clustering coefficients higher than truly random networks; however, random linguistic networks have higher clustering coefficients than true syntactic networks ($random\ networks < syntactic\ networks < randomized\ syntactic\ networks$). This is probably due to the fact that syntax has a restricting effect: for instance, "a" and "have" will never be linked in a truly syntactic network.

4.2.5. Other interesting results from syntactic network research

To end the section on syntactic networks, we present two lines of research that strengthen the idea that syntactic networks do capture properties of language.

On the one hand, Ferrer-i-Cancho, Capocci, & Caldarelli (2007) show how spectral methods of clustering can tell apart, with relative success, verbs and nouns from a syntactic network where categories are not encoded; classification for adjectives and adverbs is less effective. From these results, they optimistically conclude that "word classes could be eventually discovered using only the structure of syntactic interactions."

On the other hand, Liu & Li (2010) use seven network parameters (average degree, cluster coefficients, average path length, network centralization, diameter*, power exponent of degree distribution, and the determination coefficient of power law distributions) to classify 15 languages: Arabic, Catalan, modern Greek, ancient Greek, English, Basque, Hungarian, Italian, Japanese, Portuguese, Romanian, Spanish, Turkish, Latin, and Chinese. They use dependency-annotated corpora and achieve quite —but not completely— successful results, shown in Figure 9. It is interesting that this classification is similar to one performed taking into account only whether constructions in a text are head-initial or head-final (Liu, 2010), shown in Figure 10. It is very surprising, because word order information is lost in networks.

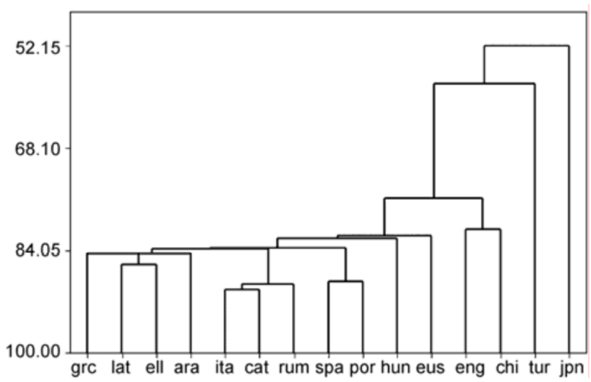

Figure 9 Liu & Li's (2010) typological classification using network parameters

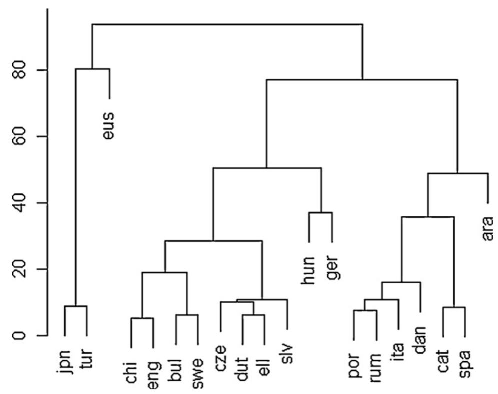

Figure 10 Liu's (2010) typological classification using word-order criteria

Later, Liu & Xu (2011) use the same parameters on lemmatized and non-lemmatized networks to carry out the same classification task, and conclude that non-lemmatized networks achieve better results.

As we have seen, the mathematical properties of syntactic graphs do seem to mirror the nature of word-to-word interactions in natural language; for instance, *C* values seem to reflect the restraining effect of syntax over word interactions, and different patterns of connectivity can be used to tell verbs and nouns apart.

However, what inferences about the cognitive basis of syntax can be drawn from syntactic networks? I believe more needs to be investigated in that area.

5. Analyzing English and Spanish networks

In this section, we will analyze networks generated from two languages and three different treebanks.

5.1. The treebanks

5.1.1. Spanish: AnCora and Universal Dependencies

AnCora-ESP (Taulé, Martí, & Recasens, 2008) is a Castilian Spanish treebank containing 500,000 word tokens and built from written news sources. In the original corpus, morphology was automatically tagged, but checked by humans at the manual syntactic annotation stage. Syntactic annotation followed phrase-structure formalisms and strived to be theory-neutral. The dependency treebank was automatically generated from the phrase-structure treebank.

The version of AnCora-ESP dependencies we will analyze is a reanalysis[5] of the original dependencies (Kolz, Badia, & Saurí, 2014). Two main differences are worth highlighting:

- In the original AnCora, n-grams that function as a lexical unit were parsed together into "multiwords." For instance, the Spanish expression "a causa de" would appear as "a_causa_de". In the new AnCora, multiwords are split into their individual units.
- In the original AnCora-ESP, the dependencies were semantically motivated, whereas the new AnCora showcases syntactically-oriented trees. A significant example of this differing approach is the treatment of auxiliary verbs. In AnCora-ESP, the governor is always the lexical verb, with the auxiliary performing the role of dependent; in the new AnCora, lexical verbs, as well as verb complements, are dependents of the auxiliary.

This second the AnCora-ESP was automatically converted from the original.

[5] Which can be downloaded at:
<https://portal.upf.edu/web/glicom/AnCora-dependency-models>

Moreover, we will also analyze a version of another Spanish corpus[6] (Mcdonald, Nivre, & Yoav, 2012), originally built as part of a multilingual project for Google and later adapted to the Universal Dependencies ("Universal Dependencies," n.d.)'s project.

The Universal Dependencies project aims at creating an annotation standard that can be applied to most languages, building the necessary framework for cross-linguistic research. It is based on Stanford dependencies (Marneffe & Manning, 2008) and Google universal part-of-speech tags (Petrov, Das, & McDonald, 2011). As the original AnCora dependency annotation, it uses semantic rather than syntactic criteria for annotation.

This treebank contains 431,587 words from different genres: newspaper articles, blogs and consumer reviews. It was automatically converted to Universal Dependencies from the original phrase-structure format.

5.1.2. English: English Web Treebank (EWT)

It was deemed adequate to analyze at least two languages in order to have the possibility of comparing results. For this purpose, the English Web Treebank (EWT)[7] (Silveira et al., 2014) was chosen.

Just as the second Spanish treebank, EWT follows the conventions of the Universal Dependencies presented above. It contains 254,830 word tokens, and was manually annotated at every level, from part-of-speech annotation to dependency structure.

As opposed to AnCora, EWT collected its material from the web, and focused on less formal documents, spanning five categories: blog posts, news-group threads, emails, product reviews, and answers from question-answer websites. For comparison purposes, two approaches could have been tackled: either maintaining one genre across corpora, so that it was guaranteed that different measures were due to the different languages, or having different genres, so that we would have two levels of comparison available: language and genre. Given the quality of EWT —being manually annotated, the second option was chosen.

5.2. Getting the corpora ready and generating the networks

Let us name the three treebanks introduced AnCora, Spanish_UD, and EWT. The three of them are lemmatized and annotated for dependencies; they differ, however, in two important aspects: part-of-speech (PoS) annotation and size.

Regarding PoS, AnCora's morphological tagging consists of an alphanumerical string of length six (maximum). The first two alphabetical characters correspond to the main morphological category and the subcategory of the token. The next characters give morphological information of gender, number, case, person, time, and mode (Taulé et al., 2008). Examples are shown in Figure 11.

[6] Which can be downloaded at:
<https://github.com/UniversalDependencies/UD_Spanish>

[7] Which can be downloaded at:
<https://github.com/UniversalDependencies/UD_English>

Word	Lemma	PoS
Si	si	CS
trabajo	trabajar	VMIP1S0
bajo	bajo	SPS00
presión	presión	NCFS000
bajo	bajar	VMIP1S0
el	el	DA0MS0
interés	interés	NCMS000
.	.	Fp

Figure 11 Morphological annotation in AnCora_ESP (from Taulé et al.(2008)*)*

Whereas the first two digits of AnCora's morphological annotation scheme yield 47 different tags, Universal Dependencies work with a reduced set of 17 tags. Therefore, it was AnCora's tags which were converted to the Universal Dependencies format. The conversion table is shown in Table 5. Some UD tags, such as PART (particle), did not apply to the languages considered and are thus not present in the table.

AnCora	UD	Notes
NC	NOUN	Common nouns
NP	PROPN	Proper nouns
P*	PRON	Pronouns, including relative pronouns.
VM, VS	VERB	Lexical verbs
VA	AUX	Auxiliary verbs
AQ	ADJ	Adjectives
R*	ADV	Adverbs
D* (except DN)	DET	Determiners
Z*, DN, AO	NUM	Numeral expressions, including determiners and ordinals.
SP	ADP	Prepositions (adpositions in Universal Dependencies, but there are no postpositions in English nor Spanish)
CC	CONJ	Coordinating conjunctions
CS	SCONJ	Subordinating conjunctions
P*	PRON	Pronouns, including relative pronouns.
F*	SYM	All symbols and punctuation marks, which were ignored when creating the networks.
(*) indicates any character		

Table 5 Part-of-speech conversion table

Regarding size, since AnCora and Spanish_UD are much larger than EWT, reduced versions of AnCora and Spanish_UD were created so that the networks generated from them displayed a similar number of nodes —in the "forms" version of the network; see below. The reduction was carried out by randomly erasing files until reaching an appropriate size.

Up to now we have described five corpora: AnCora, AnCora_half, Spanish_UD, Spanish_UD_half, and EWT. For each of them, eight different networks were generated: a simple and a directed version of each of the networks listed in Table 6. We analyzed both word form and lemma networks. Moreover, for each of them the nodes could either be represented as bare word forms, or as a string with the format "word_PoS". The distinction was necessary to be able to tell homophones apart. As an example, consider the word form "trabajo" in "Trabajo de lunes a viernes". It would appear as "trabajo" and "trabajo_VERB" in the form networks, and as "trabajar" and "trabajar_VERB" in the lemmatized networks.

AnCora_form	AnCora_half_form	Spanish_UD_form	Spanish_UD_half_form	EWT_form
AnCora_forms_PoS	AnCora__half_form_PoS	Spanish_UD_form_PoS	Spanish_UD_half_form_PoS	EWT_form_PoS
AnCora_lemma	AnCora_half_lemma	Spanish_UD_lemma	Spanish_UD_half_lemma	EWT_lemma
AnCora_lemma_PoS	AnCora_half_lemma_PoS	Spanish_UD_lemma_PoS	Spanish_UD_half_lemma_PoS	EWT_lemma_PoS

Table 6 List of networks generated in simple and directed versions.

5.3. Comparative size

Although size is apparently a very superficial measure, Liu & Xu (2011) affirmed that "difference between lemma networks and word form networks is the best criterion in language classification." This is due to the fact that that highly inflected languages will show a considerable reduction in size from their form to their lemma networks, whereas poorly inflected languages will not.

The expectation for our networks, therefore, is that the Spanish lemma networks will be much more reduced, compared to their form counterparts, than the English ones. We can observe in Table 7 (in the "lemma/form" rows) that this is in fact the case: while in Spanish networks the lemmatized version is between 65 and 78% of its non lemmatized counterpart, the English lemmatized network is only 20% smaller than the form network.

But why do AnCora and Spanish_UD differ in this measure more than 10%? In Spanish_UD, contractions, such as "del", and verb-pronoun forms such as "tomarnos" are split into two nodes: "de" and "el", "tomar" and "nos". This means, especially in the case of verbs, that variation for one lemma is considerably reduced, which would explain the smaller gap between Spanish_UD_form and Spanish_UD_lemma networks.

There are interesting differences between simple and directed graphs as well. We observe that edges grow between 1.5 and 7.5% from simple to directed networks, with edge/node correlation growing only slightly. It is natural that we observe an edge increase in directed graphs, since a relation can sometimes be established in both directions between the same two nodes; for instance, consider "broken glass" and "I have broken the glass;" according to UD, "glass" is the governor of "broken" in the first case, but the relation is reversed in the second. It is striking, however, that in AnCora and Spanish_UD, the average growth is approximately 5.8%, whereas in EWT is 2.1%. This discrepancy cannot be due to the freer word order found in Spanish, since word-order and edge direction are unrelated. This measure would indicate that, in Spanish, lexical items are more likely to function both as governors and heads, whereas in English there are fewer of such items. We will come back to this issue in section 5.5.

	SIMPLE GRAPHS			DIRECTED GRAPHS		
	Nodes (N)	Edges	Edges/N	Nodes (N)	Edges	Edges/N
AnCora						
form	39,693	210,115	5.29	39,693	221,570	5.58
form_PoS	43,911	215,772	4.91	43,911	226,331	5.15
lemma	25,666	164,673	6.42	25,666	177,065	6.90
lemma_PoS	28,509	169,519	5.95	28,509	181,349	6.36
lemma/form (%)	64.66			64.66		
form_PoS/lemma_PoS (%)	64.92			64.92		
Spanish_UD						
form	45,594	226,255	4.96	45,594	232,899	5.11
form_PoS	51,109	232,555	4.55	51,109	238,311	4.66
lemma	35,316	198,128	5.61	35,316	207,894	5.89
lemma_PoS	40,028	206,002	5.15	40,028	214,569	5.36
lemma_/form_ (%)	77.46			77.46		
form_PoS/lemma_PoS (%)	78.32			78.32		
AnCora_half						
form	19,552	79,944	4.09	19,552	83,759	4.28
form_PoS	21,265	81,638	3.84	21,265	85,149	4.00
lemma	13,283	66,355	5.00	13,283	70,761	5.33
lemma_PoS	14,529	67,934	4.68	14,529	72,110	4.96
lemma/form (%)	67.94			67.94		
form_PoS/lemma_PoS (%)	68.32			68.32		
Spanish_UD_half						
form	18,362	62,916	3.43	18,362	64,300	3.50
form_PoS	20,069	64,145	3.20	20,069	65,349	3.26
lemma	14,323	57,054	3.98	14,323	59,169	4.13
lemma_PoS	15,813	58,727	3.71	15,813	60,555	3.83
lemma/form (%)	78.00			78.00		
form_PoS/lemma_PoS (%)	78.79			78.79		
EWT						
form	18,593	132,945	7.15	18,593	135,306	7.28
form_PoS	21,851	136,181	6.23	21,851	138,003	6.32
lemma	15,510	117,029	7.55	15,510	120,594	7.78
lemma_PoS	18,477	121,780	6.59	18,477	124,526	6.74
lemma/form_(%)	83.42			83.42		
form_PoS/lemma_PoS (%)	84.56			84.56		

Table 7 Size and related measures for all networks

5.4. Main Network Measures

In Table 8 we find a summary of the main values for the simple networks of all corpora considered, which we comment one by one below.

5.4.1. Average degree of the network

The average degree of the network is the average over the degree* of all nodes. It seems that degree is directly related to two factors: the size of the network and whether vertices are word forms or lemmas. As for the former, the correlation also holds when observing Ferrer-i-Cancho et al. (2004) and Liu's (2008) data. For instance, a Chinese corpus of 16,654 words had a k value of 6.48, and second Chinese corpus of 19,960 words had an average degree of 8.91. This is not surprising, since new words are probably harder to find as a corpus grows larger, whereas connections are constantly being created due the infinite combinatory potential of language.

The fact that lemma networks have higher k values is also easy to explain, since lemma nodes group all the edges that belong to the different word form nodes.

	k	C	C_{random}	d	d_{random}
AnCora					
form	10.59	0.32	0.0002	2.92	4.74
form_PoS	9.83	0.30	0.0002	2.97	4.92
lemmatized	12.83	0.36	0.0005	2.86	4.26
lemma_PoS	11.89	0.33	0.0005	2.91	4.42
Spanish_UD					
form	9.92	0.18	0.0002	3.23	4.92
form_PoS	9.10	0.17	0.0002	3.28	5.15
lemmatized	11.22	0.25	0.0003	3.10	4.60
lemma_PoS	10.29	0.24	0.0003	3.14	4.79
AnCora_half					
form	8.18	0.27	0.0004	3.01	4.92
form_PoS	7.68	0.25	0.0003	3.06	5.11
lemmatized	9.99	0.32	0.0009	2.92	4.38
lemma_PoS	9.35	0.29	0.0005	2.97	4.54
Spanish_UD_half					
form	6.85	0.13	0.0004	3.37	5.30
form_PoS	6.39	0.12	0.0004	3.44	5.53
lemmatized	7.97	0.20	0.0006	3.18	4.86
lemma_PoS	7.43	0.18	0.0005	3.23	5.06
EWT					
form	14.30	0.14	0.0008	3.27	3.96
form_PoS	12.46	0.10	0.0005	3.42	4.24
lemmatized	15.09	0.19	0.0010	3.23	3.83
lemma_PoS	13.18	0.14	0.0007	3.36	4.07

Table 8 Representative values for all simple networks

The high value of k in the EWT network, however, is surprising. It is higher than the AnCora and Spanish_UD values, which double its size. Moreover, average path length values this high have not been found previously in the literature—not in Ferrer-i-Cancho et al. (2004), Liu (2008), nor Liu & Xu (2011). We believe these outstanding values are due to the very informal nature of the EWT corpus, completely made up from blog posts, news-group threads, emails, product reviews, and answers from question-answer websites. It is very likely that informal corpora contain fewer types than formal corpora —due to less vocabulary variation, which would translate into fewer nodes receiving more connections.

5.4.2. Clustering coefficient and average path length

All networks show small world structure, characterized by C (clustering coefficient*) values that are significantly higher than C values in random networks with the same size, and d (average path length*) values that are close to those of random networks.

Again, relations between corpus and network size are patent. If we compare AnCora and Spanish_UD to their reduced counterparts, it seems that the bigger the network, the higher its clustering coefficient.

However, taking into account that both are networks of similar size of the same language, there is a striking difference between the clustering coefficient of AnCora and Spanish_UD. We believe that this difference is due to differences in genre. AnCora spans less genres than Spanish_UD: news from only three different sources (Taulé et al., 2008) as opposed to the variety of newspaper articles, blogs and consumer reviews. The topics and vocabulary coherence of AnCora could lead to strongly connected neighborhoods that the thematically dispersed Spanish_UD might not be able to achieve.

Regarding average path length, it was calculated not on the whole networks, but on their largest component*. The reason for this is that, if there is no path between two nodes, it is considered that the distance between them is infinite; therefore, if there was only one of such pairs of nodes in the network, the average path length*—see formula in the glossary—would also become infinite. However, since language networks are characterized by high degrees of connectivity, their biggest component is usually a large fraction of the whole network, as we can see in Table 9.

The average path length in syntactic networks is similar but consistently smaller than in random networks, here, in Ferrer-i-Cancho et al. (2004) and in Liu (2008). When we look at average path length, we see the opposite correlation than for clustering coefficients: the bigger the network, the shorter its shortest path lengths.

There is, again, a striking difference between Spanish_UD and AnCora, with paths being longer in Spanish_UD despite it being larger than AnCora. We believe this effect is due to annotation differences. The purely syntactic annotation of AnCora leads some words —mainly auxiliaries— to create highly connected nodes; on the contrary, semantic annotation disperses links to lexical words. Let us see an example to clarify the explanation.

(a) I have received a present; I have run a mile; I have drunk a smoothie.

In AnCora, all verb arguments, and the main lexical verb of (a) would be dependents —would point towards— the auxiliary “have”; in Spanish_UD, on the contrary, all arguments of the verb and the auxiliary are encoded as dependents of the lexical verbs “received”, “run” and “drunk”. With just three short sentences, “have” would be at the receiving end of 7 links, whereas in Spanish_UD, “received”, “run” and “drunk” would each get 3 incoming edges.

In syntactically annotated corpora, highly connected nodes, such as auxiliaries, probably connect nodes that otherwise would be far apart, such as “smoothie” and “run”, if we consider the examples in (a). That could explain shorter d values in AnCora than in Spanish_UD; consider the directed versions of networks in Figure 12: $d_{syn}(smoothie, run) = 2$, but $d_{sem}(smoothie, run) = 3$.

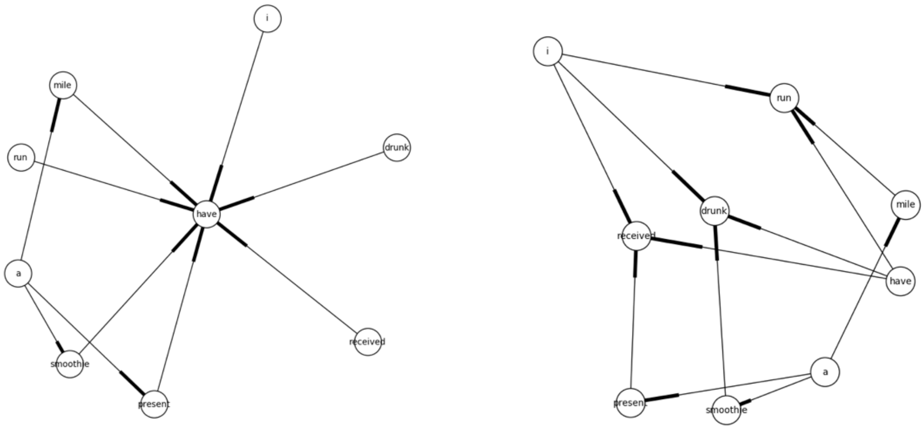

Figure 12 Syntactic versus semantic annotation and path lengths

	N_c (largest component)	N	N_c / N
Ancora			
form	39,693	39,693	1
form_PoS	43,898	43,911	0.99
lemmatized	25,664	25,666	0.99
lemmatized_PoS	28,494	28,509	0.99
Spanish_UD			
form	45,492	45,594	0.99
form_PoS	50,986	51,109	0.99
lemmatized	35,227	35,316	0.99
lemmatized_PoS	39,922	40,028	0.99
Ancora_half			
form	19,546	19,552	0.99
form_PoS	21,251	21,265	0.99
lemmatized	13,279	13,283	0.99
lemmatized_PoS	14,521	14,529	0.99
Spanish_UD_half			
form	18,312	18,362	0.99
form_PoS	20,003	20,069	0.99
lemmatized	14,275	14,323	0.99
lemmatized_PoS	15,750	15,813	0.99
EWT			
form	18,523	18,593	0.99
form_PoS	21,755	21,851	0.99
lemmatized	15,441	15,510	0.99
lemmatized_PoS	18,383	18,477	0.99

Table 9 Size of largest components for all simple networks

5.5. Power law degree distribution

As already explained, the degree distribution* $P(k)$ of a network informs us about the fraction of nodes that have degree* k. Scale free* degree distributions* follow the formula:

$$P(k) \approx k^{-\gamma}$$

This means that edges with many vertices are scarce, while edges with few vertices are very abundant. If plotted on log-log axes, scale free degree distributions show a straight line with a long tail, as we can see in Figures 13 and 14, representing AnCora_form_PoS and EWT_form_PoS (similar patterns were found for all networks).

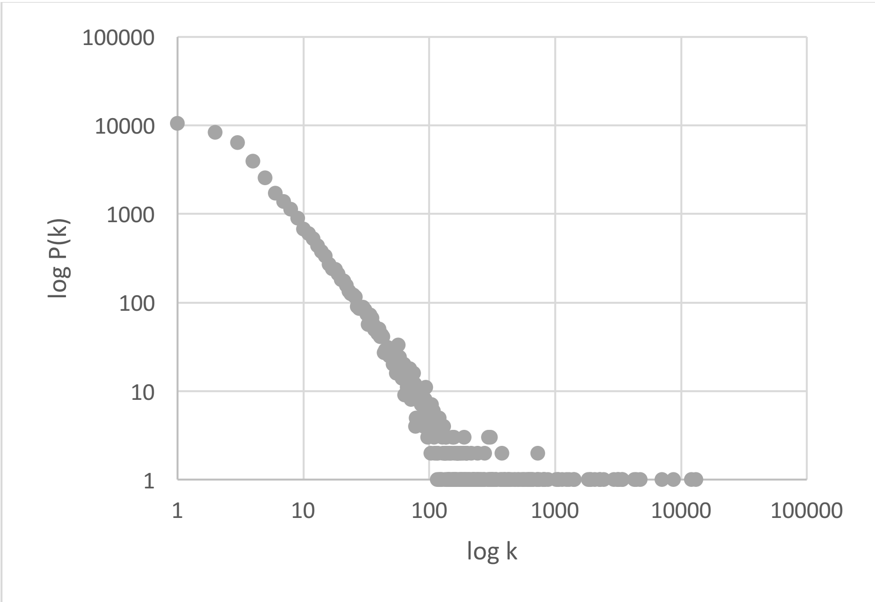

Figure 13 Log/log degree distribution for AnCora_form_PoS

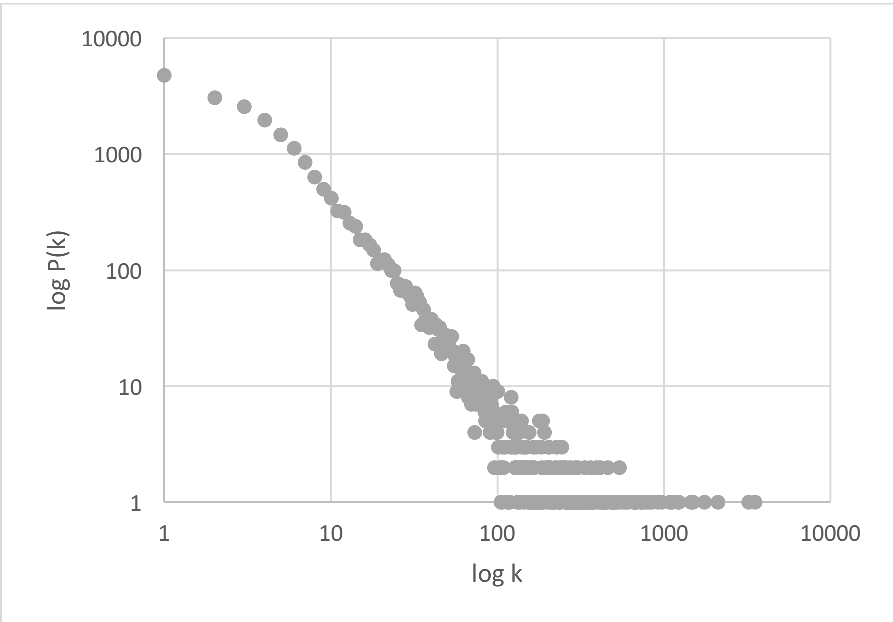

Figure 14 Log/log degree distribution of EWT_form_PoS

Just as in Ferrer-i-Cancho et al. (2004) and in Liu (2008), we find that the scale free degree distributions for all networks have an exponent than ranges between 1.80 and 2.50 (Table 10).

	SIMPLE GRAPHS	DIRECTED GRAPHS	
	γ	γ_{in}	γ_{out}
AnCora			
form	2.39	2.42	2.28
form_PoS	2.37	2.39	2.30
lemmatized	2.36	2.30	2.39
lemma_PoS	2.35	2.29	2.33
Spanish_UD			
form	2.00	2.10	1.97
form_PoS	2.05	2.18	2.01
lemmatized	1.92	1.97	1.94
lemma_PoS	1.96	2.02	1.91
AnCora_half			
form	2.37	2.44	2.28
form_PoS	2.34	2.33	2.32
lemmatized	2.35	2.30	2.36
lemma_PoS	2.30	2.28	2.36
Spanish_UD_half			
form	2.27	2.48	2.15
form_PoS	2.32	2.55	2.20
lemmatized	2.06	2.66	2.06
lemma_PoS	2.11	2.64	2.11
EWT			
form	2.31	2.82	2.09
form_PoS	2.30	2.74	2.10
lemmatized	1.85	2.68	2.20
lemma_PoS	2.43	2.71	1.82

Table 10 Exponent of degree distribution for all networks

But to further understand the structure of the language networks, it is crucial to know which are the scarce but highly connected vertices. In what follows we show the 20 most connected nodes for the lemma_PoS networks —hypothesizing that effects will be stronger on lemma than on form networks.

In simple networks (Table 11), just as Ferrer-i-Cancho et al. (2004) had pointed out, the first places of the list are occupied by function words, although very frequent verbs are also present.

In directed networks, we can differentiate between in-degree* and out-degree*. Since nodes point to their governors, out-degree (Table 12) is a measure of the frequency and combinatorial capacity of a word; words positioning high need to appear often and be linked to different governors. The same relation appearing 10 times in a corpus will only be reflected in the network as an edge connecting two nodes; this means that if lemma el_DET has 7,957 outgoing edges, it must have been combined with different words that many times in the corpus.

	AnCora_lemma_PoS		Spanish_UD_PoS		EWT_PoS	
	node	k	node	k	node	k
1	de_ADP	10,277	de_ADP	9,707	the_DET	3,078
2	y_CONJ	8,291	y_CONJ	5,290	be_VERB	2,748
3	el_DET	8,162	en_ADP	3,991	i_PRON	901
4	en_ADP	5,460	uno_DET	3,425	make_VERB	836
5	a_ADP	4,417	a_ADP	3,317	be_AUX	822
6	ser_VERB	3,793	él_PRON	3,022	do_VERB	715
7	del_ADP	3,769	ser_VERB	2,922	or_CONJ	637
8	por_ADP	3,380	con_ADP	2,050	have_AUX	615
9	uno_DET	3,319	por_ADP	1,917	come_VERB	562
10	con_ADP	3,179	su_DET	1,709	see_VERB	525
11	haber_AUX	2,830	tener_VERB	1,638	by_ADP	474
12	para_ADP	2,325	hacer_VERB	1,094	but_CONJ	471
13	al_ADP	2,073	haber_AUX	1,041	one_NUM	454
14	su_DET	1,617	estar_VERB	861	put_VERB	392
15	o_CONJ	1,593	año_NOUN	828	year_NOUN	387
16	como_SCONJ	1,551	o_CONJ	814	‘s_ADV	362
17	estar_VERB	1,442	este_DET	803	this_DET	360
18	tener_VERB	1,409	no_ADV	801	try_VERB	355
19	que_PRON	1,307	encontrar_VERB	763	some_DET	353
20	que_SCONJ	1,283	más_ADV	736	this_PRON	313

Table 11 Nodes with highest k in simple lemma_PoS networks

We observe that, in AnCora, high positioning lemmas are functional words, determiners and prepositions. In position 17 and 20 we find two verbs: “haber_AUX” and “ser_VERB”. We expect “haber_AUX” to be a governor with incoming, rather than outgoing edges (as we will see in Table 13). However, in subordinate clauses the main verb—in AnCora’s case, the auxiliary— points to the conjunction that introduces the subordinate, which can explain “haber” and a verb as frequent as “ser” in the first positions of the out-degree list.

In Spanish_UD, again, functional words top the list, along with some very common verbs: “ser_VERB”, “haber_AUX”, “tener_VERB”, and “hacer_VERB”. Again, this is probably due to their presence in subordinate phrases—in the case of UD, they are dependents of the head verb of the main clause. We can observe that for the Spanish networks, the lists are extremely similar.

As for EWT’s out-degree list, we find again functional words, and some very common verbs such as “make_VERB” or “do_VERB.”

If dependents point to their governors, in-degree (Table 13) measures how often a word functions as a syntactic or semantic head. Here we expect lexical words to appear more often and, indeed, we find more verbs and some nouns in the lists.

In Ancora, as expected, verbs are auxiliaries and those verbs that can form part of periphrastic constructions: “ser_VERB”, “haber_AUX”, “estar_VERB”, “tener_VERB”, y “poder_VERB”. In Spanish_UD, on the other hand, lexical verbs: “tener_VERB”, “hacer_VERB”, “encontrar_VERB”, “llegar_VERB”, “realizar_VERB”[8] and some frequent nouns, such as “parte” and “año”.

	AnCora_lemma_PoS		Spanish_UD_lemma_PoS		EWT_lemma_PoS	
	node	k_{out}	node	k_{out}	node	k_{out}
1	el_DET	7957	de_ADP	8573	the_DET	3071
2	de_ADP	5599	el_DET	7446	be_VERB	2235
3	uno_DET	3155	y_CONJ	4812	i_PRON	847
4	en_ADP	2913	en_ADP	3598	be_AUX	810
5	del_ADP	2323	a_ADP	2705	or_CONJ	636
6	y_CONJ	2146	uno_DET	2625	have_AUX	606
7	por_ADP	1847	ser_VERB	2065	by_ADP	473
8	a_ADP	1794	con_ADP	1802	but_CONJ	468
9	su_DET	1592	él_PRON	1768	‘s_ADV	362
10	con_ADP	1416	por_ADP	1697	this_DET	359
11	para_ADP	1302	su_DET	1433	this_PRON	295
12	que_PRON	1090	o_CONJ	757	some_DET	287
13	al_ADP	1021	este_DET	753	do_AUX	273
14	él_PRON	991	haber_AUX	682	one_NUM	270
15	más_ADV	859	no_ADV	595	about_ADP	263
16	como_SCONJ	804	más_ADV	573	make_VERB	232
17	haber_AUX	706	año_NOUN	520	do_VERB	227
18	este_DET	704	tener_VERB	458	year_NOUN	210
19	no_ADV	687	también_ADV	419	very_ADV	204
20	ser_VERB	624	hacer_VERB	400	only_ADV	199

Table 12 Nodes with highest out-degree for directed lemma_PoS networks

In EWT the list is dominated by very frequent verbs, and some frequent nouns that are very similar to those found in Spanish_UD: “year” and “group”.

Note, however, a striking difference between Spanish and English lists: in Spanish, functional words are in high ranking positions of both in- and out-degree lists, whereas in EWT out-degree is almost fully composed of verbs. If you remember, when analyzing the size of networks, we observed that the edge-increase from simple to directed networks was larger for Spanish than English. Here we find an explanation: it seems that prepositions are tagged as not only as dependents as would be expected, but also governors more often in Spanish than in English. Is this a matter of language typology or annotation practices?

The difference between semantically and syntactically oriented annotation is most patent in the out-degree list, a significant list containing the most important heads in a given corpus.

[8] It must be noted that, while in AnCora the only verb marked as AUX is “haber”, in UD all verbs participating in periphrastic constructions are tagged as AUX.

	AnCora_lemma_PoS		Spanish_UD_lemma_PoS		EWT_lemma_PoS	
	node	k_{in}	node	k_{in}	node	k_{in}
1	de_ADP	7808	de_ADP	2191	be_VERB	924
2	y_CONJ	7684	él_PRON	1706	make_VERB	671
3	ser_VERB	3554	tener_VERB	1332	do_VERB	567
4	en_ADP	3389	ser_VERB	1272	come_VERB	450
5	a_ADP	3326	uno_DET	1112	see_VERB	424
6	haber_AUX	2527	a_ADP	956	put_VERB	328
7	con_ADP	2077	y_CONJ	853	try_VERB	297
8	del_ADP	2011	hacer_VERB	813	ask_VERB	239
9	por_ADP	1912	en_ADP	700	feel_VERB	230
10	para_ADP	1446	encontrar_VERB	617	one_NUM	216
11	o_CONJ	1335	estar_VERB	580	year_NOUN	197
12	al_ADP	1278	uno_PRON	512	group_NOUN	184
13	estar_VERB	1269	haber_AUX	433	pay_VERB	174
14	tener_VERB	1231	llegar_VERB	430	become_VERB	166
15	que_SCONJ	1114	haber_VERB	428	recommend_VERB	154
16	como_SCONJ	952	decir_VERB	415	experience_NOUN	148
17	poder_VERB	950	realizar_VERB	409	stay_VERB	145
18	hacer_VERB	803	parte_NOUN	391	check_VERB	142
19	deber_VERB	551	su_DET	377	read_VERB	139
20	e_CONJ	536	año_NOUN	369	mean_VERB	136

Table 13 Nodes with highest in-degree for directed lemma_PoS networks

5.6. Hierarchical organization

Hierarchical organization of a network (Ravasz & Barabási, 2002) is identified through k and C values: we can affirm that this feature is present when degree and clustering coefficient are inversely correlated. We do find that our networks are hierarchically organized, as the Figures 15 and 16 show (similar patterns were found for all networks).

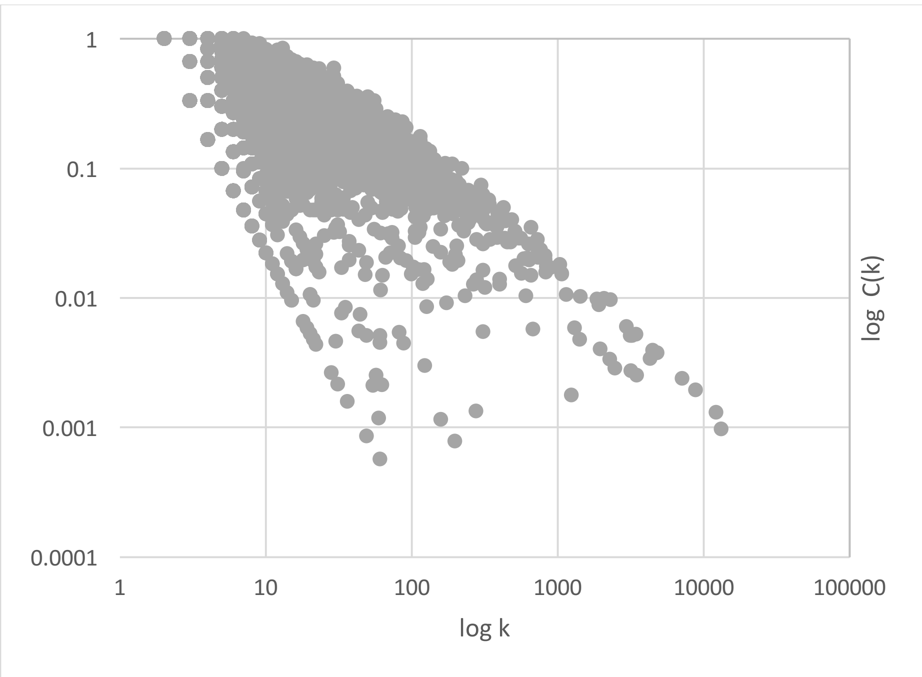

Figure 15 Log/log k/C(k) correlation for Ancora_form_PoS (directed)

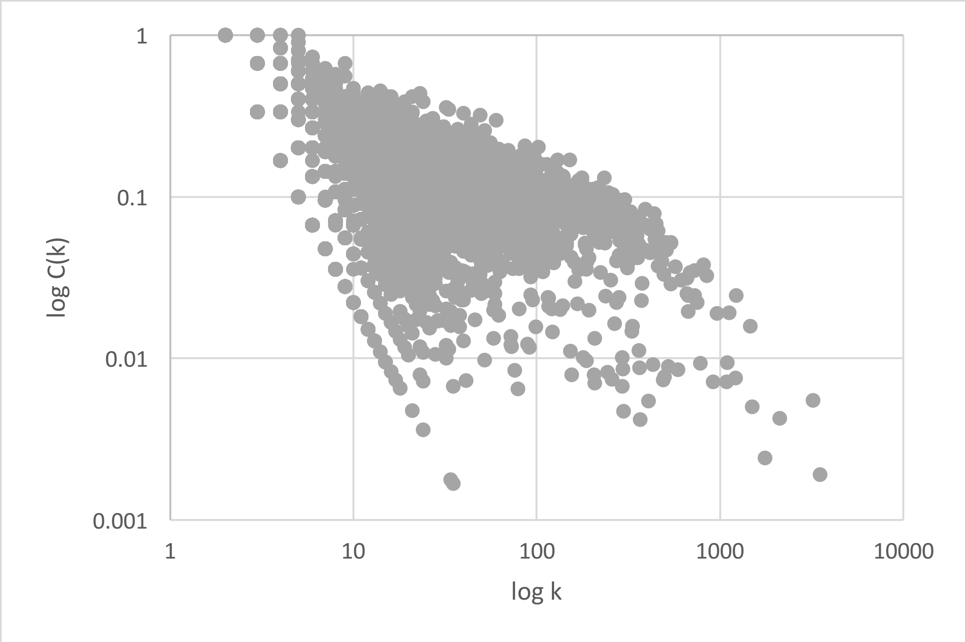

Figure 16 Log/log k/C(k) correlation for EWT_form_PoS (directed)

Ferrer-i-Cancho et al. (2004) assumed that hierarchy in syntactic networks was related to the hierarchic nature of syntax. We mentioned that, if that was the case, nodes with highest k and lowest C values should be word forms that usually occupy head positions in sentence structure, such as verbs, which are usually the root —head— of sentences. However, we consistently find that the list of strongly connected nodes is topped by closed class words such as prepositions and determiners; this points to word frequency as a determining factor for hierarchical structure in syntactic networks.

In directed networks, however, we did find many verbs in the out-degree ranked lists, especially for EWT. Do these nodes also have low clustering coefficients? In Ancora_lemma_PoS, "ser_VERB" and "haber_AUX" appear in position 20 and 28 in a list ordered according to the clustering coefficient of each word in ascending order (once all nodes with a C value of 0 have been removed). In Spanish_UD_lemma_PoS, "tener_VERB," "hacer_VERB," "encontrar_VERB," "llegar_VERB," and "realizar_VERB" occupy positions 39, 91, 98, 263 and 228. Although these are not the highest positions, they are relatively high if we consider the size of the networks.

So it might be true that, in part, syntactical hierarchy is reflected as network hierarchy as Ferrer-i-Cancho et al. (2004) suggested; nevertheless, word frequency seems to play a stronger part in that role.

6. Collecting our thoughts

In this Master's thesis we have looked at semantic and syntactic network research. The former rests on a body of previous research in linguistics —relational semantics— and psycholinguistics —the spreading activation model, priming effects, etcetera— which seem to have laid a foundation from which semantic network research could successfully take off. Its syntactic counterpart, on the other hand, has had very little handed down to it. Even the most basic question —what is it that syntactic networks represent? — is still unanswered (and, worryingly, not even asked often).

We have observed that some regularities of language translate to certain values in syntactic network properties such as clustering coefficient, average path length or degree distribution. This is not surprising. However, we have also seen that changes in treebank size, annotation paradigm, and text genre affect network properties, too. Being aware of the effect that extralinguistic factors have on networks makes us wonder about the extent of their influence. Are some of the results we are getting an artifact of the annotation paradigm?

Dependencies were chosen for their straightforward implementation and the fact that, as opposed to mere co-occurrence patterns, they reflect syntactic or semantic relations as defined by the linguistics community. In fact, there is no alternative at the moment that could substitute dependencies in syntactic network generation. However, the lack of choices does not mean that the option available can fulfill our needs. We must ask whether dependencies are successful at capturing what we are trying to capture: patterns of interaction among words in real language use.

But, what kind of interaction are we interested in? Lexical units interact in more than one way; the fact that there are syntactic and semantic dependency annotation standards reflects this fact. Words in a sentence are able to establish *semantic relationships* thanks to the structure that grammar provides, that is, thanks to *syntactic relations*.

If we are merely interested in semantic relations, perhaps doing away with functional words would be the best option; ignore articles, particles, prepositions and even auxiliaries, and create networks with lemmas as nodes and edges coded for relation type. It would not be hard to build such networks from semantic dependency treebanks such as the ones being created under the Universal Dependencies project. This method would yield networks in which semantically related items form closely-knitted neighborhoods and polysemic words play the role of connecting them. This type of representation, however, would have to be studied as a dynamic semantic network, never as a syntactic network.

On the other hand, if what we are trying to get is an insight of syntactic relations, semantically annotated treebanks should not be used as source material[9]. But do the current syntactic dependencies suffice?

What syntactic dependencies try to capture is the *structure* hidden behind the linear presentation of linguistic content. According to Melčuk (1988), therefore, the means by which this structure is accomplished —linear order, prosody and inflections— are irrelevant. Syntactic annotation should always be lemmatized and relevant relations reflected through typed links.

> It seems fairly obvious that syntactic means, i.e., devices used by natural languages to encode syntactic structure in actual sentences, cannot be part of the structure itself: otherwise, we have a flagrant *contradictio in terminis*. Therefore, syntactic word order,

[9] At present it is hard to know what kinds of treebanks are being used, since papers exploring linguistic networks do not mention the type of dependency of their source treebanks.

prosody and inflections should be banned from the representation of syntactic structure (Melčuk, 1988)

If we follow this line of reasoning, it would seem that functional words should also be ignored in syntactic networks that aim to capture structure, since their function is exactly the same as linear order, prosody and inflections: to give rise to structure. Moreover, if we do include functional words, a deep asymmetry is created cross-linguistically. For instance, let us take the case of a language with a rich case system, such as Finnish, and a language that relies mainly on prepositions, such as Spanish. If we only ignore inflections, the lemmatized networks of Finish will only display a few functional words, whereas in Spanish these will be extremely abundant and highly connected. In fact, it could very well be that the typological classification carried out by Liu & Li (2010) and Liu & Xu (2011) was an artifact of differing syntactic annotation standards for different languages.

Alternatively, it could be the case that a researcher is interested precisely in the different means by which languages accomplish structural organization. In that case, functional words and morphemes would play a very important role, but also word order, which dependencies disregard. Perhaps this line of research is impossible to carry out through networks, or perhaps we could devise means to make it possible. For instance, the network could be two layered, in the sense that one layer of edges represents syntactic relations, whereas the second layer of directed edges represents word order.

Furthermore, an essential aspect of dependency annotation is that syntactic relations are typed. Current networks treat all links homogeneously for practical reasons: methods of analysis for networks with typed links —multilayer networks— are not as developed, although efforts are being made in that direction (Kivelä et al., 2014), even specifically in the language area (Martinčić-Ipšić, Margan, & Meštrović, 2016).

Not only edges, but nodes are also a matter of concern. We saw that the original AnCora contained multiwords, and that in the Universal Dependencies' corpus of Spanish contractions such as "del" or "dormirse" were split. The debate over the classification of words is never-ending, and it is very likely that there is a continuum in this category. Are "del" and "mirarlo" words? Is "a causa de" a word? Are frequent pairs such as "cat nap" or "copy cat" words? Whichever notion of word one decides to use, the decision should be made consciously and not be limited to whatever is available, especially if we are trying to draw conclusions about cognitive principles.

The concerns outlined above are not extremely deep, and begs the question: why are these simple reflections not taken into account in current research? If we have a look at language network research, we will notice that most of current language network research is carried out by physicists and published in physics journals. Although finding the relevant mathematical properties of language networks or developing new ways to analyze them is certainly a job that only physicians and mathematicians can carry out, it is also true that the linguistic knowledge they display in their papers is far from deep. If linguists took part in modelling the networks and in relating mathematical results to grammatical properties, it would be possible to get a clearer view of *what* we are analyzing and *how* to interpret the mathematical properties of syntactic networks.

GLOSSARY

Average path length ($\boldsymbol{d}$)

- Of a word (d_v). If $d_{min}(i,j)$ is the minimum path* connecting vertices i and j, $d_v(i)$ is the sum of the d_{min} from i to every other node in the network, divided by N, the size* of the network.

$$d_v(i) = \frac{1}{N}\sum_{j=1}^{N} d_{\min}(i,j)$$

- Of a network (d). The sum of the d_v of every node in the network, divided by N.

$$d = \frac{1}{N}\sum_{i=1}^{N} d_{min}(i)$$

Average path length measures how many steps it takes, on average, to go from any node in the network to any other.

Betweenness Centrality ($\boldsymbol{g}$)

Number of shortest paths among all nodes that go through vertex *v.*

$$g(v_i) = \sum_{st} n_{st}^{i}$$

Where n_{st}^{i}corresponds to 1 if v_i lies on the geodesic path* between v_s and v_t, and 0 if it does not. According to this measure, in undirected networks geodesic paths are counted twice (st, ts). $s = t$ cases are usually excluded.

Bipartite network

A bipartite network contains two sets of nodes. The nodes in each set can only be linked to nodes in the other set, but never to nodes in the set they belong to.

Clustering coefficient ($\boldsymbol{C}$)

- Of a node (C_v). The fraction of neighbors* of a node that are also neighbors of each other. That is, of triples* including v, how many are transitive*?

$$C_v = \frac{(number\ of\ triangles *\ including\ v_i)}{(number\ of\ triples *\ including\ v_i)}$$

- Of a network (C). The sum of C_v for all nodes in the network, divided by the size* of the network.

$$C = \frac{1}{N}\sum_{i=1}^{N} C_v(v_i)$$

C measures the average fraction of pairs of neighbors of a node that are also neighbors of each other.

Component

Subset of vertices of a network such that it is always possible in it to find a path from any vertex to any other vertex.

Degree of a node (k)

- In a simple network, the degree of a vertex $k(v)$ is the number of edges in the network that are linked to v.
- In a directed network we can distinguish between:
    - In-degree of a vertex $k_{in}(v)$, the number of incoming edges to v from other nodes in the network.
    - Out-degree of a vertex $k_{out}(v)$, the number of outcoming edges from v to other nodes in the network.

Degree distribution ($\boldsymbol{P(k)}$)

The fraction of vertices that have degree k. If $p_0, p_1, p_2, \ldots$ represent the number of vertices with 0, 1, 2… edges attached to them, and N is the size* of the network, then:

$$P(k) = \frac{p_k}{N}$$

Diameter (D)

The length of the longest path between any pair of vertices in the network for which a path actually exists .

Eccentricity

The eccentricity of a node *v* is the maximum distance from *v* to all other nodes in G.

Geodesic path

Shortest path* connecting two nodes in a network

Largest connected component

See *connected component**.

Neighbors

Two nodes sharing an edge are said to be neighbors.

Path

A walk* from an origin edge to a destination edge that fulfills the following conditions:

- No repeated vertices
- No repeated edges

Power law degree distribution

Also called scale free degree distribution. Degree distributions that vary as a function of node degree* are called power laws. Degree distribution in scale free networks follows the formula:

$$P(k) \approx k^{-\gamma}$$

Random graph

A graph initiated with a specified N number of nodes in which edges between them are established randomly.

Size of a network

Number of vertices it contains.

Scale free*

See *power law degree distribution**.

Small world network

A network displaying the following features:

- Average path length* (d) is approximately the same as the d of a random network of the same size*.
- Clustering coefficient* (C) is highter than the C value of a random network of the same size*.

Transitivity

The property that, if A has relation R with B, and B has relation R with C, then A has relation R with C. In network theory, the relation we are concerned with is "sharing an edge."

Triangle

A triple* where each of the nodes is connected to the other two nodes by an edge.

Triple

Three connected vertices.

Walk

A succession of edges, such that consecutive edges share an edge: $(v_i, v_j), (v_j, v_k), \ldots$

Weighted network

Network where edges have a value associated to them, indicating the strength of the connection.

References

Barabási, A.L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512. http://doi.org/10.1126/science.286.5439.509

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*(4), 343–355. http://doi.org/10.1016/0013-4694(85)90008-2

Bhalla, U. S., & Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science (New York, N.Y.)*, *283*(5400), 381–7.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., … Wiener, J. (2000). Graph structure in the Web. *Computer Networks*, *33*(1-6), 309–320. http://doi.org/10.1016/S1389-1286(00)00083-9

Brown, G., & Watson, F. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, *15*(3), 208–216. http://doi.org/10.3758/BF03197718

Brown, R. (1958). How shall a thing be called. *Psychological Review*, *65*(1), 14–21. http://doi.org/10.1037/h0041727

Bybee, J. (2010). *Language as a compex adaptive system: the interaction of cognition, culture and use*. *Language, Usage and Cognition*. New York: Cambridge University Press. http://doi.org/10.1017/CBO9781107415324.004

Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller, & M. Halle (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Carey, S. (1985). Constraints on semantic development. In J. Mehler (Ed.), *Neonate Cognition* (pp. 381–398). Hillsdale (N.J.): L. Erlbaum Associates.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, *25*(1), 85–95. http://doi.org/10.1080/14640747308400325

Chomsky, N. (1965). Aspects of the Theory of Syntax. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press. http://doi.org/10.1016/0732-118X(86)90008-5

Choudhury, M., & Nukherjee, A. (2009). The structure and dynamics of linguistic networks. In N. Ganguly, A. Deutsch, & A. Mukherjee (Eds.), *Dynamics on and of complex networks: applications to biology, computer science, and the social sciences* (pp. 145–166). Boston: Birkhäuser.

Collins, A. M., & Quillian, M. R. (1969). Retrieval Time from Semantic Memory. *Journal of Verbal Learning and Behavior*, *8*, 240–247.

Cong, J., & Liu, H. (2014). Approaching human language with complex networks. *Physics of Life Reviews*, *11*(4), 598–618. http://doi.org/10.1016/j.plrev.2014.04.004

Ellis, N. C., & Larsen-Freeman, D. (2006). Language Emergence: Implications for Applied Linguistics - Introduction to the Special Issue. *Applied Linguistics*, *27*(4), 558–589. http://doi.org/10.1093/applin/aml028

Elman, J. L. (1995). Language as a dynamical system. In R. F. Port & T. van Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (pp. 195–223). Cambridge, MA: MIT Press.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*(7), 301–306. http://doi.org/10.1016/j.tics.2004.05.003

Érdi, P. (2008a). Complex Systems: The Intellectual Landscape. In *Complexity Explained* (pp. 1–20). Berlin: Springer.

Érdi, P. (2008b). History of Complex Systems Research. In *Complexity Explained* (pp. 25–55). Berlin: Springer.

Estrada, E. (2011). *The structure of complex networks: theory and applications*. New York: Oxford University Press.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database.* Cambridge, MA: MIT Press.

Ferrer-i-Cancho, R. (2006). Why do syntactic links not cross? *Europhysics Letters (EPL)*, *76*(6), 1228–1235. http://doi.org/10.1209/epl/i2006-10406-0

Ferrer-i-Cancho, R. (2014). Beyond description. *Physics of Life Reviews*, *11*(4), 621–623. http://doi.org/10.1016/j.plrev.2014.07.014

Ferrer-i-Cancho, R., Capocci, A., & Caldarelli, G. (2007). Spectral methods cluster words of the same class in a syntactic dependency network. *International Journal of Bifurcation and Chaos*, *17*(07), 2453–2463. http://doi.org/10.1142/S021812740701852X

Ferrer-i-Cancho, R., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B*, *268*(1482), 2261–2265. http://doi.org/10.1098/rspb.2001.1800

Ferrer-i-Cancho, R., & Solé, R. V. (2001). Two Regimes in the Frequency of Words and the Origins of Complex Lexicons: Zipf's Law Revisited. *Journal of Quantitative Linguistics*, *8*(3), 165–173. http://doi.org/10.1076/jqul.8.3.165.4101

Ferrer-i-Cancho, R., Solé, R. V, & Köhler, R. (2004). Patterns in syntactic dependency networks. *Physical Review E*, *69*(5), 051915. http://doi.org/10.1103/PhysRevE.69.051915

Freeman, W. (2000). *Mesoscopic Brain Dynamics*. London: Springer-Verlag.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-Acquisition, Word Frequency, and Neighborhood Density Effects on Spoken Word Recognition by Children and Adults. *Journal of Memory and Language*, *45*(3), 468–492. http://doi.org/10.1006/jmla.2000.2784

Geeraerts, D. (2010). Relational semantics. In *Theories of Lexical Semantics* (pp. 74–90). Oxford, New York: Oxford University Press.

Gilhooly, K. J., & Logie, R. H. (1981). Word age-of-acquisition, reading latencies and auditory recognition. *Current Psychological Research*, *1*(3-4), 251–262. http://doi.org/10.1007/BF03186735

Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford, New York: Oxford University Press.

Hopper, P. (1987). Emergent Grammar. In *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics* (pp. 139–157).

Hudson, R. (2007). *Language Networks: The New Word Grammar*. Oxford: Oxford University Press. http://doi.org/10.1017/CBO9781107415324.004

Hudson, R. (2014). The view from linguistics. Comment on "Approaching human language with complex networks" by Cong and Liu. *Physics of Life Reviews*, *11*(4), 619–620. http://doi.org/10.1016/j.plrev.2014.06.013

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813. http://doi.org/10.3758/BF03196544

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. http://doi.org/10.1073/pnas.0707835105

Kivelä, M., Barthelemy, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, (July), 203–271. http://doi.org/10.1093/comnet/cnu016

Kohn, K. W. (1999). Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molecular Biology of the Cell*, *10*(8), 2703–34.

Kolz, B., Badia, T., & Saurí, R. (2014). From constituents to syntax-oriented dependencies. *Procesamiento Del Lenguaje Natural*, (52), 53–60.

Krishna, M., Hassan, A., Liu, Y., & Radev, D. (2011). *The effect of linguistic constraints on the large scale organization of language* (Arxiv preprint No. arXiv:1102.2831v2).

Kucera, H., & Francis, W. N. (1968). Computational Analysis of Present-Day American English. *American Documentation*, *19*(4), 419. http://doi.org/10.2307/302397

Larsen-Freeman, D. (1997). Chaos I Complexity Science and Second Language Acquisition. *Applied Linguistic*, *18*(2), 144–165.

Liu, H. (2008). The complexity of Chinese syntactic dependency networks. *Physica A: Statistical Mechanics and Its Applications*, *387*(12), 3048–3058. http://doi.org/10.1016/j.physa.2008.01.069

Liu, H. (2010). Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua*, *120*(6), 1567–1578. http://doi.org/10.1016/j.lingua.2009.10.001

Liu, H., & Hu, F. (2008). What role does syntax play in a language network? *EPL (Europhysics Letters)*, *83*(1), 18002. http://doi.org/10.1209/0295-5075/83/18002

Liu, H. T., & Li, W. W. (2010). Language clusters based on linguistic complex networks. *Chinese Science Bulletin*, *55*(30), 3458–3465. http://doi.org/10.1007/s11434-010-4114-3

Liu, H., & Xu, C. (2011). Can syntactic networks indicate morphological complexity of a language? *EPL (Europhysics Letters)*, *93*(2), 28005. http://doi.org/10.1209/0295-5075/93/28005

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618–630. http://doi.org/10.3758/BF03212999

Marneffe, M. De, & Manning, C. D. (2008). Stanford typed dependencies manual. http://doi.org/10.1.1.180.3691

Martinčić-Ipšić, S., Margan, D., & Meštrović, A. (2016). Multilayer network of language: A unified framework for structural analysis of linguistic subsystems. *Physica A: Statistical Mechanics and Its Applications*, *457*, 117–128. http://doi.org/10.1016/j.physa.2016.03.082

Mcdonald, R., Nivre, J., & Yoav, Y. Q. (2012). Universal Dependency Annotation for Multilingual Parsing. *ACL*, (2), 92–97.

Mehler, A. (2015). Towards a theoretical framework for analyzing complex linguistic networks.

Melčuk, I. A. (1988). *Dependency Syntax: Theory and Practice*. New York: State University of New York Press.

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & Cognition*, *25*(1), 47–56. http://doi.org/10.3758/BF03197284

Miller, G. A. (1995). WordNet : A Lexical Database for English, *38*(11), 39–41.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., … Alon, U. (2004). Superfamilies of evolved and designed networks. *Science (New York, NY)*, *303*(5663), 1538–1542. http://doi.org/10.1126/science.1089167

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of Acquisition Norms for a Large Set of Object Names and Their Relation to Adult Estimates and Other Variables. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 528–559. http://doi.org/10.1080/027249897392017

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida Word Association, Rhyme and Fragment Norms.

Newman, M. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404–409. http://doi.org/10.1073/pnas.98.2.404

Newman, M. (2010). *Networks: an introduction*. Oxford, New York: Oxford University Press.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, *96*(14), 8028–8033. http://doi.org/10.1073/pnas.96.14.8028

Nowak, M. A., Plotkin, J. B., & Jansen, V. A. A. (2000). The evolution of syntactic communication. *Nature*, *404*(6777), 495–498. http://doi.org/10.1038/35006635

Oltvai, Z. N., Barabási, A.-L., Jeong, H., Tombor, B., & Albert, R. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651–654. http://doi.org/10.1038/35036627

Petrov, S., Das, D., & McDonald, R. (2011). A Universal Part-of-Speech Tagset.

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410–430. http://doi.org/10.1002/bs.3830120511

Ravasz, E., & Barabási, A. (2002). Hierarchical Organization in Complex Networks. *Physical Review E*, *67*(2), 026112. http://doi.org/10.1103/PhysRevE.67.026112

Saussure, F. de. (1972). Cours de linguistique générale. *Collection Payothèque*.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. http://doi.org/10.1037/0033-295X.96.4.523

Sigman, M., & Cecchi, G. A. (2002). Global organization of the Wordnet lexicon. *Proceedings of the National Academy of Sciences*, *99*(3), 1742–1747. http://doi.org/10.1073/pnas.022341799

Silveira, N., Dozat, T., De Marneffe, M.-C., Bowman, S., Connor, M., Bauer, J., & Manning, C. (2014). A Gold Standard Dependency Corpus for English. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 2897–2904.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214–241. http://doi.org/10.1037/h0036351

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences*, *102*(33), 11629–11634. http://doi.org/10.1073/pnas.0409746102

Steels, L. (2000). Language as a complex adaptive system. In *International Conference on Parallel Problem Solving from Nature* (pp. 17–26). Berlin: Springer.

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, *29*(1), 41–78. http://doi.org/10.1207/s15516709cog2901_3

Strogatz, S. H., & Watts, D. J. (1998). Collective Dynamics of "Small-World" Networks. *Nature*, *393*(June), 440–442.

Taulé, M., Martí, M. A., & Recasens, M. (2008). *Ancora: Multilingual and multilevel annotated corpora*. (ELRA, Ed.)*LREC, Marrakech*. Morocco.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.

Tomasello, M. (2003). *Constructing a language : a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition*, *26*(6), 1282–1291. http://doi.org/10.3758/BF03201200

Universal Dependencies. (n.d.). Retrieved August 2, 2016, from http://universaldependencies.org/

Williams, R. J., & Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, *404*(6774), 180–183. http://doi.org/10.1038/35004572