The Structure and Dynamics of Linguistic Networks

Monojit Choudhury[1] and Animesh Mukherjee[2]

[1] Microsoft Research India, Sadashivnagar, Bagalore – 560080
monojitc@microsoft.com
[2] Department of Computer Science and Engineering,
Indian Institute of Technology, Kharagpur, India – 721302
animeshm@cse.iitkgp.ernet.in

1 Introduction

Human beings as a species is quite unique to this biological world for they are the only organisms known to be capable of thinking, communicating and preserving potentially an infinite number of ideas that form the pillars of modern civilization. This unique ability is a consequence of the complex and powerful human languages characterized by their *recursive syntax* and *compositional semantics* [40]. It has been argued that language is a dynamic complex adaptive system that has evolved through the process of self-organization to serve the purpose of human communication needs [80]. The complexity of human languages have always attracted the attention of physicists, who have tried to explain several linguistic phenomena through models of physical systems (see e.g., [32, 42]).

Like any physical system, a linguistic system (i.e., a language) can be viewed from three different perspectives [52]. On one extreme, a language is a collection of *utterances* that are produced by the speakers of a linguistic community during the course of their interactions with other speakers of the same community. This is analogous to the *microscopic* view of a thermodynamic system, where every utterance and its corresponding context contributes to the identity of the language, i.e., the grammar. On the other extreme, a language can be characterized by a set of grammar rules and a vocabulary. This is analogous to a *macroscopic* view. Sandwiched between these two extremes, one can also conceive of a *mesoscopic* view of language, where linguistic entities, such as the letters, words or phrases are the basic units and the grammar is an emergent property of the interactions among them.

Complex networks provide a suitable framework to model and study the structure and dynamics of linguistic systems from a mesoscopic perspective. Although multi-agent simulation is the preferred modeling paradigm for microscopic studies in linguistics (see e.g., [15, 80]), there has been some work where networks are also involved. For instance, in [67], the interaction pattern between the agents are modeled as a social network and the diffusion of linguistic innovations (which are key to language change) are studied on various network topologies. This survey is confined to the works pertaining to various linguistic networks only at the level of mesoscopy.

There has been a plethora of work on linguistic networks with various motivations and at various levels of linguistic structure. On the basis of the primary goal of the research, the work in this area can be broadly classified into two categories: (1) those which intend to investigate the structural properties of language from the perspective of language evolution and thereby, explain the emergence of certain universal characteristics of languages, and (2) those which try to exploit the network based representations to develop certain useful practical systems such as machine translation, information retrieval and summarization systems. This article focuses on the former works, but a brief overview of the latter is also presented in Sec. 5.

The survey is organized from the perspective of linguistic structure. Sec. 2 describes lexical networks, where the nodes are words and edges represent lexical relationship between two words such as phonetic and semantic similarity. In Sec. 3 we present an overview of various networks where again the nodes are the words, but unlike the case of lexical networks, the edges represent their co-occurrences in similar context. These networks are representations of the interactions among words as governed by the grammar rules of a language. Sec. 4 describes the phonological networks, where the nodes are sublexical units such as phonemes or syllables. Applications of linguistic networks in Natural Language Processing (NLP) and Information Retrieval (IR) are discussed in Sec. 5. Sec. 6 concludes the survey by enumerating some open problems in the area of linguistic networks.

2 Lexical Networks

The phrase "mental lexicon" (henceforth ML) usually refers to the repository of word forms that is assumed to reside in the human brain. The average size of the receptive vocabulary for a normal high school student has been found to be more than 100,000 [63]. Quite surprisingly, speakers are capable of navigating this huge lexicon in a very efficient way; reaction time to judge whether a word form is legitimate takes less than 100 milliseconds. Consequently, there can be two important questions associated with ML: (a) how are the words stored in the long term memory, i.e., how ML is organized, and (b) how are these words retrieved from ML. Note that, the above questions are highly inter-related – to predict the organization one can investigate how words are retrieved from ML and vice versa.

One of the earliest attempts to model the organization of ML has been made in [13]. In this work, the authors propose a hierarchical structure of ML, where the concepts are arranged in the form of a tree and the attributes of a particular concept in this tree can be inherited by all the child-concepts. Figure 1 shows a representative example formed from the concepts "animal", "mammal", and "fish". While early studies as the above focused mainly on representation of the local structure of ML, its global structure remained largely unexplored.

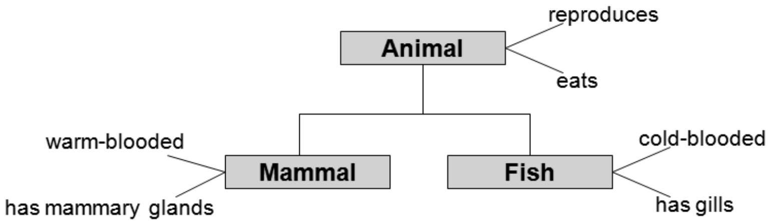

Fig. 1. The Hierarchical Structure of ML.

Recently, researchers have also started to investigate the global structure of ML primarily within the framework of complex systems and more specifically complex networks (see [36, 45, 77, 83, 86] for reference). In all of these studies ML is modeled as a web of inter-connected nodes, where each node corresponds to a word form and the inter-connections may be based on any one (or more) of the following:

- Phonological similarity (e.g., the words banana, bear, and bean may be connected since they start with the same phoneme),
- Semantic similarity (e.g., the words banana, apple, and pear may be connected since all of them are names of fruits),
- Frequency of usage,
- Age at which the word forms are acquired,
- Parts-of-speech, and
- Orthographic properties.

In the rest of this section we review one representative study each (referring, wherever applicable, to the other relevant ones) of such complex networks constructed based on (a) phonological, (b) semantic, and (c) orthographic similarities of the word forms. Syntactic similarity based networks will be discussed in detail in the next section.

2.1 Phonological Similarity based Networks

Phonological similarity among the word forms has been extensively studied in the past to infer the structure of ML and consequently, the nature of a linguistic system [4, 35, 71, 81]. This large-scale phonological ML has also been studied in the framework of complex networks in which the word forms represent the nodes and two nodes (read words) are connected by an edge if they differ only by the addition, deletion or substitution of one or more phonemes [36, 45, 83, 86]. [45] reports one of the most popular studies, where the author constructs a *Phonological Neighborhood Network* (PNN) in order to unfurl the organizing principles of ML. In PNN there is an edge (u, v) connecting the nodes u and v iff at least $\frac{2}{3}$ of the phonemes that occur in the word represented by u also occur in the word represented by v. For instance, if the word is 6 phonemes long, then one can derive all its neighbors by changing at most two phonemes through insertions, deletions, and substitutions.

The author uses the Hoosier Mental Lexicon database [68] and builds the above network from the phonologically transcribed forms of each word present in the database. More specifically, he constructs a directed network, where a long word can have a short word as its neighbor without the short word being the neighbor of the long word. For instance, if the number of segments in which the two words, say w_1 and w_2, differ is less than $\frac{1}{3}$ of the length of w_1, then there will be a directed edge from the node corresponding to w_1 to the node corresponding to w_2. The fraction $\frac{1}{3}$ is chosen, because it has been useful in earlier experiments for predicting reaction times and familiarity ratings (see [53] for reference).

The author shows that PNN is characterized by a very high clustering coefficient (0.235) but at the same time exhibits a long average path length (6.06) and diameter (20). This indicates that, like a small-world network, the lexicon has many densely inter-connected neighborhoods. However, links between two nodes from different neighborhoods are harder to find unlike small-world networks.

Low mean path lengths are necessary in networks that are to be traversed quickly, the purpose of traversal being search in most of the cases. However, in the case of ML, the search should not inhibit the neighbors of the stimulus' neighbors that are non-neighbors of the stimulus itself and are therefore, not similar to the stimulus. Hence, it can be conjectured that, in order to search in PNN, traversal of links between distant nodes is usually not required. In contrast, the search involves an activation of the structured neighborhood that share a single sub-lexical chunk that could be acoustically related during word recognition [55].

Further, the author shows that the degree distribution of the nodes in PNN is exponential rather than scale-free. Thus, one can posit that the structure of ML is not consistent with "growth via preferential attachment" at least for the neighborhood density metrics used for this study. This is because, for the standard preferential attachment model, the emergent degree distribution of the network is known to be scale-free [5]. The cause for the emergence of the exponential degree distribution for PNN is not yet well understood and is quite an open area for further research.

2.2 Semantic Similarity based Networks

One of the classic examples of semantic similarity based networks is the Wordnet [20]. In this network, concepts (known as synsets) are the nodes and semantic relationships between them are represented through the edges. In [77] the authors analyze the structure of the nouns in the English Wordnet database (version 1.6). The semantic relationships between the nouns can be primarily of four types (i) hypernymy/hyponymy (e.g., animal/cat), (ii) antonymy (e.g., day/night), (iii) meronymy/holonymy (e.g., trunk/tree), and (iv) polysemy (e.g., the concepts "the main stem of a tree", "the body excluding the head and neck and limbs", "a long flexible snout as of an elephant" and "luggage consisting of a large strong case used when travelling or for storage" are connected to each other due to the polysemous word "trunk" which can mean all of these). Some of the important findings of this work are:

- Semantic relationships are scale invariant,
- The hypernymy tree forms the skeleton of the network,
- Inclusion of polysemy reorganizes the network into a small-world,
- The nodes with the most traffic (i.e., nodes with maximum number of paths passing through them) correspond to those concepts, which are expressed by the most polysemous words. They are also found to have very high clustering coefficient,
- In presence of polysemous edges, the distance between two nodes across the network is not in correspondence with the depth at which they are found in the hypernymy tree.

Further references to the studies on such semantic relationship based networks can be found in [1, 82]. Although there are several work attempting to analyze the structure of the semantic network of words, one hardly finds any study explaining the emergence of these topological properties through models of network synthesis. It would be very interesting to study the correlates of semantic acquisition and symbol grounding with the model parameters.

2.3 Orthographic Similarity based Networks

Like phonological similarity networks, one can also construct networks based on orthographic similarity, where the nodes are the words and the edit-distance between two words defines the edge weight between the nodes corresponding to them. Such networks have been studied in order to investigate the difficulties involved in spelling error detection and correction [11]. In this work the authors construct such networks (SpellNet) for three different languages (Bengali, Hindi and English) and analyze them to show that

- For a particular language, the probability of real word errors can be equated to the average weighted degree of SpellNet,
- The hardness of non-word error correction correlates to the average clustering coefficient for a language,

- The basic topological properties are invariant in nature for all the languages; for instance the authors find that the SpellNet for all of the three languages are characterized by an exponential degree distribution, high clustering coefficent and positive correlation between the degree and clustering coefficient of the nodes.

3 Word Co-occurrence Networks

In this section, we review the work on *word co-occurrence networks* where the nodes are the words and an edge between two words indicate that the words have co-occurred in the language in certain context(s). Depending on the definition of the context, various networks can be defined. We describe in detail two such networks, namely the collocation network and the syntactic dependency network. As an application, we discuss the work by [79] where the collocation network has been used for unsupervised induction of the grammatical structure of a language.

3.1 Collocation Network

One of the most basic and well studied co-occurrence networks are the *word collocation networks*, where two words are linked if they are neighbors, that is they collocate, in a sentence [24]. In this work, two types of collocation networks – the *unrestricted* and the *restricted* ones – were constructed for English from The British National Corpus. In the unrestricted network, all the collocation edges are preserved, whereas in the restricted one only those edges are preserved for which the probability of occurrence of the edge is higher than the case when the two words collocate independently. All these networks are undirected and unweighted, even though in language order of the words ("ticket book" is different from "book ticket") as well as the frequency of the collocations have obvious significance.

The authors found that both the networks exhibit small world properties. The average path length between any two nodes is small (around 2 to 3) and the clustering coefficients are high (0.69 for the unrestricted and 0.44 for the restricted networks). However, the most striking observation regarding these networks is that the degree distributions follow a two regime power-law. The degree distribution of the 5000 most connected words follow a power-law with an exponent -3.07, which is surprisingly close to that of the Barabási-Albert growth model [5]. These findings led the authors to argue that the word usage of the human languages is preferential in nature, where the frequency of a word defines the comprehensibility and production capability. Thus, the higher the usage frequency of a word, the higher is the probability that the speakers will be able to produce it easily and the listeners will comprehend it fast. This is known as the *recency effect* in linguistics [3]. The small-world property of the collocation network on the other hand makes it easier to search the mental lexicon. In essence, the authors conclude that evolution of language has resulted in an optimal structure of the word interactions that facilitate easier and faster production, perception and navigation of the words.

It does not follow, however, from the collocation networks that a word with high degree is indeed a word with high usage frequency (unless the word co-occurrences are completely independent in nature, which essentially is not the case). In a separate study, Cancho and Solé [25] have shown that the rank-degree distribution of the words in a very large corpus also follows a two regime power-law supporting their claim regarding the presence of a core-lexicon whose size is about 5000 words. In order to explain the two regime power-law in word collocation networks, Dorogovtsev and Mendes [18] proposed a preferential attachment based growth model. At every time step t, a new word (i.e., a node) enters the language (i.e., the network) and connects itself preferentially to one of the pre-existing nodes. Simultaneously, ct (where c is a positive constant) new edges are grown between pairs of old nodes that are chosen preferentially. Through mathematical analysis and simulations, the authors establish that this model gives rise to a two regime power-law with exponents very close to those observed in [24].

There have been studies on the properties of collocation networks for languages other than English. Examples include Russian [46] and many others [41]. The basic topological properties of the networks (e.g., scale-free, small world, assortative) are similar across languages and thus, points to the fact that like Zipf's law, these characteristics are also linguistic universals and call for a non-trivial psycho-linguistic account of their emergence and existence.

3.2 Syntactic Dependency Network

Although collocation networks are easier to construct, they do not necessarily capture the syntactic and semantic relationships between the words. This is because syntactic and semantic relations often extend beyond the local neighborhood of a word. Syntactic relations between the words of a language are governed by the underlying grammar. There are various formalisms, such as phrase structure grammar, tree-adjoining grammar and dependency grammar, to capture these relationships. In the dependency grammar formalism a relationship, often shown as a directed edge, connects two words – the *head* and the *dependent*. The dependent word modifies the head word in a certain way. For example, the nouns are the heads of the adjectives that modify them. Similarly, the verbs are the heads of their subjects, objects and other arguments. Thus, in the dependency formalism, every sentence is represented as a directed acyclic graph or the dependency tree as illustrated in Fig. 2. Usually, the finite verb is the head of the whole sentence and is not dependent on any other word.

Cancho and his co-authors [21, 26] defined the *syntactic dependency network* (SDN) where the words are the nodes and there is a directed edge between two words if in any of the sentences of a given corpus there is a directed

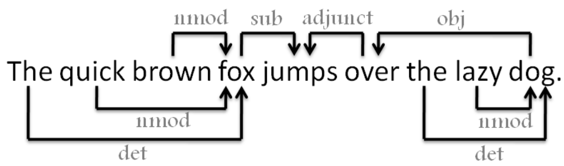

Fig. 2. Example of a dependency tree. The arrows are labelled by the type of depedency relation and run from the dependant to the head words.

dependency relation between these words. The direction of the dependencies in their construction is from the dependent word to the head word. In order to construct the SDN, one needs to know the dependency relations between the words of a sentence. Fortunately, there are large dependency treebanks for some languages consisting of human annotated dependency trees for a several thousand sentences. The authors studied the SDN for three languages: Czech, German and Romanian, and observed strikingly similar characteristics.

All the networks exhibit power-law degree distributions and small world structure. Some of the very interesting topological properties observed are:

- *Disassortative mixing*: This shows that words that are used for linking other words (such as prepositions) and therefore, have high degree in the networks, are not linked themselves.
- *Hierarchical organization*: It implies that there is a top down hierarchy that is the basis of phrase structure formalism.
- *Small world structure*: This is necessary for recursion and fast navigation of the mental lexicon.

It is a well known fact that syntactic dependency links usually do not intersect in any of the world's languages. In [22], the author conjectured that this phenomenon is an outcome of minimization of the Euclidian distance between the syntactically related words of a sentence, where the Euclidian distance between two words is given by the number of words separating them[3]. Later on, Cancho et al [23] showed that spectral clustering of SDN puts words belonging to the same syntactic categories in the same cluster. As we shall see in Sec. 5, quite similar techniques are being used in the field of NLP for unsupervised induction of syntactic categories.

3.3 Unsupervised Grammar Induction

One of the fascinating applications of word collocation networks, illustrated in [79], is related to unsupervised induction of grammar. Explaining the pro-

[3] While it is true that syntactic dependencies have a tendency to avoid crossing, there are systematic exceptions to that generalization in languages with relatively free constituent order. In German, for example, about one third of all relative clauses are extraposed, thus creating crossing dependencies.

cess of language acquisition is one of the greatest challenges to modern science. Children learn languages, that they are exposed to, quite accurately and effortlessly. This is one of the strongest evidence in support of our instinctive capacities towards languages [70], which is dubbed as the Universal Grammar by Noam Chomsky [10]. In [79], the authors proposed a very simple algorithm for learning hierarchical structures from the collocation graph of a raw text corpus. The algorithm, ADIOS, works as follows.

A directed collocation graph is constructed from the corpus, where the words are the nodes and an edge is drawn from words w to v if v follows w in a sentence. In fact, each sentence is represented as a separate path in the graph. The algorithm then iteratively searches for *motifs* that are shared by different sentences. A linguistic motif is defined as a sequence of words, which tends to occur quite frequently in the language and also serves some special function. For example, "that the X is Y" is a very commonly occurring motif in English, where X and Y can be substituted by a large number of words and this whole pattern can be embedded in various parts of a sentence.

Solan et al. [79] define the probability of a particular structure being a motif in terms of network flows. After finding the motifs, the algorithm proceeds to identify interchangeable motifs and merge them into a single node. Thus, at every step the network becomes smaller and a hierarchical structure emerges. This structure can then be presented as a set of phrase structure grammar rules.

ADIOS has a high precision ($\approx 70\%$), but low recall ($\approx 40\%$). Through a comparative analysis of the induced grammars, the authors were able to construct a dendogram of 6 languages that have been studied. Quite surprisingly, the dendogram reflects the phylogenetic relations between these 6 languages. There are other graph based methods for unsupervised induction of syntactic structures, but unlike ADIOS, these algorithms are based on standard probability theory and Bayesian models.

4 Phonological Networks

In the earlier sections, we have seen how complex networks can be used to study the different types of interactions (phonological, syntactic, and semantic) between the words of a language. In this section, we shall review some of the work, where the networks are constructed from linguistic units that are smaller than words, e.g., phonemes and syllables.

4.1 Network of Human Speech Sounds

The most basic units of human languages are the speech sounds. The repertoire of sounds that make up the sound inventory of a language are not chosen arbitrarily, even though the speakers are capable of perceiving and producing a plethora of them. In contrast, the inventories show exceptionally regular patterns across the languages of the world, which is arguably an outcome of the self-organization that goes on in shaping their structure. In fact, a large number of computational models have been proposed in the literature in order to explain the self-organization of the vowel inventories [15, 47, 51, 76]. A few attempts have also been made in the area of linguistics to reason the observed patterns across the consonant inventories. Most of these works confine themselves to explaining certain individual principles rather than formulating a general theory describing the emergence of them. However, complex networks have been recently used quite successfully to explain the self-organization of the consonant inventories. In [65] the authors construct a bipartite network called PlaNet or the Phoneme-Language Network in which one of the partitions consist of nodes representing the languages while the other partition consists of nodes representing the consonants. There is an edge between the nodes of these two partitions if a particular consonant occurs in a particular language. The authors further construct PhoNet (Phoneme-phoneme Network), which is the one-mode projection of PlaNet onto the consonant nodes i.e., a network of consonants in which the nodes are linked as many times as they have co-occurred across the language inventories. The data used for constructing the above networks is drawn from the UCLA Phonological Segment Inventory Database (UPSID) [54] that consists of 317 languages and 541 consonants that are found across these languages. The authors observe that,

From the study of PlaNet [65]

- The degree distribution of the consonant nodes in PlaNet roughly follows a power law with an exponential cut-off towards the tail,
- A synthesis model based on preferential attachment (a language node attaches itself to a consonant node depending on the current degree (k) of the consoant node) can explain the emergence of the degree distribution of PlaNet. The results match the empirical data more accurately if the attachment kernel is super-linear (i.e., the attachment probability is proportional to k^{α}, where $\alpha > 1$).

From the study of PhoNet [64, 65]

- The degree distribution of the consonant nodes in PhoNet also roughly indicate a power-law behavior with exponential cut-offs,
- The clustering coefficient of PhoNet (=0.89) is significantly higher than that of a random graph with the same number of nodes and edges (=0.08),
- Community structure analysis of PhoNet can capture the strong patterns of co-occurrence of consonants that are prevalent across the languages of the world,
- The driving force that leads to the emergence of these communities is feature economy, which states that languages tend to use a small number of *distinctive* features and maximize their combinatorial possibilities to generate a large number of consonants,
- The emergence of the degree distribution and the clustering coefficient of PhoNet can be explained through a synthesis model that is based on both

preferential attachment and *triad* (i.e., fully-connected triplet) formation. While the preferential part of the model reproduces the degree distribution of the network, the triad formation part imposes a large number triangles onto the generated network thereby increasing the clustering coefficient,
- The emergence of feature economy can be explained by having a synthesis model, which is a linear combination of two different parts, one being driven by the usual degree-dependent preference and the other by a factor that favors the choice of those consonants that share a lot of features with the already chosen ones.

The authors postulate that the physical significance of the synthesis models are grounded in the process of language change. Language change is a collective phenomenon that functions at the level of a population of speakers [80]. They also conjecture that it is possible to explain the significance of the models at the level of an individual, primarily in terms of the process of language acquistion. Further, they argue that there are two orthognal preferences: (a) the occurrence frequency of a consonant, and (b) the feature-dependent preference (that increases the ease of learning), which are instrumental in the acquisition of the inventories. The synthesis model is essentially a linear combination of these two mutually orthogonal factors.

4.2 Network of Syllables

The syllable inventory of each language can also be modeled and analyzed in the framework of a complex network. Each node in this network is a syllable and links are established between two syllables each time they are shared by a word. In [78] the authors report the study of the network of Portuguese syllables from two different sources: a Portuguese dictionary (DIC) and the complete work of a very popular Brazilian writer – Machado de Assis (MA). The authors show that

- The networks have a low average shortest path (DIC: 2.44, MA: 2.61),
- The networks indicate a high clustering coefficient (DIC: 0.65, MA: 0.50),
- Both the networks show a power-law behavior.

Since in Portuguese the syllables are close to the basic phonetic units unlike the case in English, the authors argue that the properties of the English syllabic network should be different from that of Portuguese. The authors further conjecture that since Italian has a strong parallelism between its structure and syllable hyphenization it is possible that the Italian syllabic network has properties close to that of the Portuguese network pointing to certain universal characteristics of language.

5 Applications in NLP and IR

Graph based approaches are quite common in the areas of NLP and IR. Interestingly, although there are no obvious technical differences between the scope of graph theory in these areas and in complex networks, the terminologies used and the objectives are often quite different. The works on linguistic networks discussed in the last three sections were primarily targeted to the statistical physics community and the objective was to unfurl the structure of languages and their dynamics. In this section, we will survey some equally interesting and significant works, which use the same set of mathematical tools, but the objective is to develop practical applications concerning languages.

5.1 Induction of Syntactic and Semantic Categories

One of the earliest and recurrent applications of networks in NLP have been in automatic induction of syntactic and semantic categories based on the *distributional hypothesis* [39]. The distributional hypothesis states that words of similar syntactic (semantic) category are found in similar contexts [39]. To illustrate this concept, consider two unknown words **X** and **Y** that occur in the following sentences:

(1) The red **X** is very beautiful.

(2) If you **Y** then I shall punish you.

Even though we do not know what **X** and **Y** are, it is easy to infer that the former is a *noun* and the latter is a *verb*. We can draw such inferences about the syntactic categories (in this case the parts-of-speech) of words based on our knowledge that nouns, but not verbs, can be preceded by articles (*the*) and adjectives (*red*). The concept of distributional hypothesis is equally relevant for semantic categories. Words belonging to the same domain club together. Thus, the word *student* is expected to be in vicinity of the word *school*, rather than *market*.

Measuring to what extent two words appear in similar contexts measures their similarity [62]. The general methodology [12, 27, 31, 72, 74, 75] for inducing word class information can be outlined as follows:

1. Define the context of a word as a vector. It could be just the set of words which occur in the same sentence, or only the immediate neighbors of the words. For syntactic class induction, usually the word order is preserved during construction of the vectors and the context vectors are defined only in terms of the function words (such as *is, of, the* and *a*).
2. Collect global context vectors for the words by summing up the local contexts.
3. Construct a weighted network, where the nodes are the words and the weight of the edge between two words is the *distance* between their context vectors. There are several ways to define the distance between the vectors. Some of the common measures are Euclidian distance, cosine similarity and correlation coefficients.
4. Apply a clustering algorithm on these networks to obtain the word classes.

In the syntactic category induction literature, the 150-250 words with the highest frequency are considered as function words and the context vectors are defined based on them. Some authors employ a much larger number of features and reduce the dimensions of the resulting matrix using Singular Value Decomposition [72, 74]. [27] use the Spearman Rank Correlation Coefficient and a hierarchical clustering, [74, 75] uses the cosine between vector angles and Buckshot clustering, [31] use cosine on Mutual Information vectors for hierarchical agglomerative clustering and [12] applies Kullback-Leibler divergence in his CDC algorithm.

[28] does not sum up the contexts of each word in a context vector, but uses the most frequent instances of four-word windows in a co-clustering algorithm [16]: rows and columns (here words and contexts) are clustered simultaneously. Two-step clustering is undertaken by [74]): clusters from the first step are used as features in the second step. More recently, Biemann [6] proposed the Chinese Whispers algorithm for clustering, which is fast and does not require any parameters to be specified. [7] reports application of Chinese Whispers for POS induction in English, Finnish and German, very recently which has also been applied to Bengali [66]. In this work, the authors also investigate the topological properties of the word networks so constructed and report a scale-free degree distribution, high CC and power-law cluster size distribution.

Widdows and Dorow [87] propose an unsupervised incremental cluster building approach for acquisition of semantic classes. There are also graph based algorithms to infer semantic classes (sets of synonyms, to be specific) from the lexicons (see e.g., [17, 43]).

Identification of syntactic or semantic classes is of great importance to NLP and IR. For instance, parts-of-speech (POS) tagging is the first step towards parsing. However, the supervised machine learning techniques for POS tagging demand a large amount of human annotated data which is expensive as well as non-existent for most of the languages. Since automatic induction of POS-tags through graph clustering do not require annotated data, it might turn out to be a very useful technique in NLP for resource poor languages. Similarly, semantic clustering of the words is useful for search and IR.

5.2 Word Sense Disambiguation

Word sense disambiguation (WSD) refers to the task of assigning the appropriate sense or meaning to a word in a given context (i.e. sentence or paragraph) out of the several possibilities. For example, the English word *bank* has two different meanings as a noun: 1) river bank, and 2) a financial institution. However, as shown in the following sentences, in a given context only one of the senses is appropriate.

(1) They were walking down the bank enjoying the cool river breeze.

(2) She went to the bank to encash her cheque.

There are several ways in which graph based techniques have been applied for WSD. Examples include lexical chaining [29], semantic relatedness measures based on path lengths and random walks on semantic network [57, 61] and lexicon graphs [50]. Due to the paucity of space, here we discuss in detail

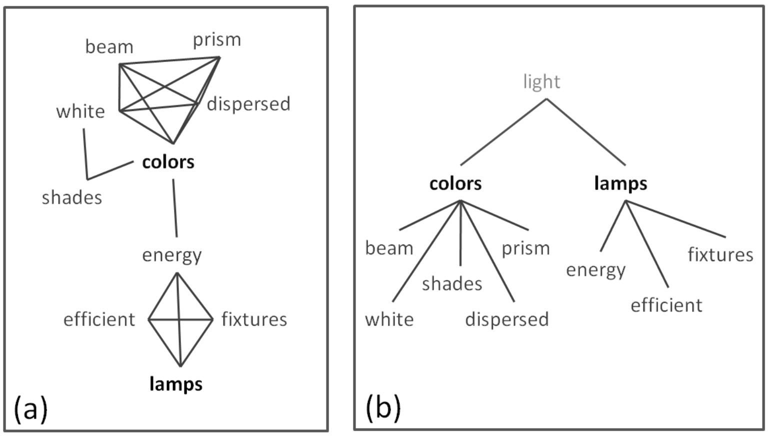

Fig. 3. Example of Hyperlex: (a) the network of words for disambiguation of the word "light"; (b) the minimal spanning tree obtained after introduction of thw word "light". The hubs are shown in bold font.

only one of the approaches – HyperLex [85] that rely on the word co-occurrence graphs.

Consider the problem of automatically identifying and disambiguating the various senses of the word *light*. The HyperLex algorithm works as follows. A subcorpus consisting of all the paragraphs featuring at least one occurrence of the word *light* is extracted from a raw text corpus. A word co-occurrence graph is constructed from this subcorpus, where the nodes are the content words except the word *light*. Two words are connected by an edge if they co-occur in a paragraph more than a preset number of times. The weight of an edge decreases as the number of times the words co-occur increases. It has been found that word co-occurrence graphs built in this manner exhibit small world properties.

In this co-occurrence network, nodes with very high degree are identified as *hubs*. The word *light*, for which we want to build the disambiguator, is then introduced to the network and connected to the hubs. A minimal spanning tree is constructed from the co-occurrence graph, where *light* is the root node and the first level consists of the hubs. Fig 3 illustrates this process. Each node in the spanning tree can be thought of as a sense. Thus, the hubs denote the basic senses and as we move further down the tree, we have more refined senses of the word. This tree can then be used for disambiguating the sense of the target word (here *light*) in a particular context.

5.3 Information Retrieval

The central problem of IR is to rank a given collection of documents with their similarity to a query. Queries are usually very short and the collection of documents huge. In a typical IR setup, the whole web consisting of billions of webpages represents this collection of documents to be ranked and the query is only one or two words long. One of the challenges of IR is to utilize the network structure of the web to compute the ranks of the documents. The web can be conceptualized as a directed graph where the nodes are the webpages and a hyperlink from webpage A to webpage B represents a directed edge between the nodes corresponding to A and B.

PageRank [9] is one of the first and very popular ranking algorithms that is allegedly used by Google search engine. The basic idea behind the PageRank algorithm is – the rank (or popularity) of a node is a function of the rank of its neighbors. In other words, the page which has a hyperlink from a popular page is also popular. An alternative view of the PageRank algorithm involves a *random walker* (here a random surfer). A random walker starts from a random node and follows the edges of the graph randomly to reach other nodes. The PageRank of a page is proportionate to the probability that a random surfer reaches that page by following random hyperlinks on the web. Yet another way to define PageRank is that it is the components of the principal eigenvector of the nodes. Thus, PageRank is also known as *eigenvector centrality* in the complex network literature.

PageRank considers only the incoming edges of a node. Kleinberg [48] proposed another ranking algorithm, called HITS, where every node has two scores – *hub* and *authority*. The *authority* scores are similar to PageRank, whereas the *hub* scores are based on the outgoing links, but computed in the same way. The final rank of a node is the combination of its hub and authority scores. Kleinberg and co-authors [33] also demonstrated how eigenvectors of the web structure can be used to cluster and disambiguate the pages corresponding to ambiguous words such as "Jaguar" (referring to an animal or a football team or the car).

One drawback of both PageRank and HITS is that the algorithms assume that all the hyperlinks have the same importance. There are various modifications of these algorithms, which use machine learning techniques to learn weights of the different types of hyperlinks. Examples include RankNet [73], TrustRank [37] and NetRank [2]. Link analysis, as this field is popularly referred to, is a very active area of research in the IR community. Some of the other emerging applications of complex networks in IR include mining social networks and blogs. The Blogosphere [49], for example, can be represented as a multi-tier network, where blogs, bloggers and other webpages (typically news articles) are the nodes and there are various types of edges representing the social network of bloggers, the links between blogs, and those between the blogs and other webpages. Analysis of the Blogosphere network is useful in classification and personalized suggestion of blogs, opinion and sentiment analysis, as well as investigating the dynamics of the world of blogs.

5.4 Other Applications

Due to paucity of space, it is impossible to do justice to the network based techniques in NLP and IR. There are a variety of NLP tasks ranging from parsing to text summarization, where graph based methods have been applied. In the previous three subsections we have discussed three specific problems to illustrate the various usages of such techniques. Before we wrap up this section, we list a few more example applications to give a flavor of the extent and potential of graph based techniques in these areas.

Text summarization is one of the notably important as well as challenging applications of NLP, which has been elegantly modeled within the framework of complex networks. The problem of text summarization involves identification of a small number of sentences from a set of given documents that best summarize the content of the documents. In [19] summarization has been reformulated as the problem of finding out the *node-centrality* in a network whose nodes are the sentences and the edges represent the word-level similarity between two sentences. The most central sentences are those which cover most of the ideas present in the given documents.

Other application areas include dependency parsing [56], textual entailment [38], sentiment classification [34, 69], keyword extraction [60], novelty detection [30] and prepositional phrase disambiguation [84]. See [8, 58, 59] for further references.

6 Conclusion

So far we have seen that there has been quite a substantial amount of work to understand the structure and dynamics of languages at the mesoscopic level within the framework of complex networks. A parallel thread of research in the field of NLP and IR tries to achieve a different goal, but uses very much the same means. Nevertheless, mesoscopic models of language as well as network based approaches to NLP are in a nascent state, especially when compared to similar lines of research in the fields of biology, economics and other social sciences (refer to the surveys in this volume). On the other hand, there seems to be a great potential for application of complex network theory to a variety of open problems in linguistics and language engineering.

One of the fundamental problems of linguistics is characterization and explanation of *linguistic universals*, i.e., properties that are common to all human languages. Differences among the languages, on the other hand, are restricted by the *typologies* and *implicational hierarchies* [14]. We have seen that, like Zipf's law, there are many linguistic universals observable in the linguistic networks. For example, the SDNs as well as word collocation networks of all languages exhibit scale-free degree distributions and small world property. A systematic investigation of topological universals of linguistic networks can substantially improve our understanding of languages. At the same time, there are properties for which the linguistic networks vary across languages. For example, the average degree of the SpellNets are very different for English, when compared to Hindi or Bengali. This difference has been attributed to the different writing systems used by English (which is alphabetic) and the two Indo-Aryan languages (which is abugida). Typological variations have also been predicted in the topological properties of syllable networks. Thus, it would be interesting to have a typological theory of languages based on the structure of the linguistic networks.

Another question of great importance for any linguistic network is about the emergence of its structural properties. It is least clear why should the word collocation networks display small world and scale-free properties. Even though the Dorogovtsev and Mendes model [18] can explain the emergence of two-regime power law observed in the collocation networks, it does not explain by itself the validity and the physical significance of this model based on preferential attachment. In other words, the phenomenon of preferential attachement at the mesoscopic level needs an independent microscopic explanation in terms of psycholinguistic factors, because words cannot voluntarily link to other words. Similar microscopic explanations are required for the non-trivial topological properties of the other linguistic networks, such as ML, SDN, PhoNet and SpellNet. This is presumably a hard problem, but any mesoscopic explanation is incomplete without a corresponding microscopic model.

In the context of NLP and IR applications, network based models are mostly ad hoc and this reduces their credibility and thereby, the popularity, as compared to the more principled Bayesian approaches. A network based language model can bridge this gap and provide us with a more systematic way of solving the NLP problems within this framework. Although there has been some initiatives in this direction [44], this area is largely unexplored and presents numerous challenging problems. Another relatively unexplored, but potentially fecund, area of research is processes "on" linguistic networks. Navigation of the ML can be modeled as guided random walks on the ML network; similarly, typographical errors can be modeled as walks on SpellNet. The exact nature of such guided walks are still to be explored and can provide strong understanding of underlying cognitive principles.

In the previous sections we have seen several ways to define networks where the nodes represent words. One can conceive of a universal word network obtained through superimposition of these partial representations of a linguistic system into a multi-tier network where the nodes are the words and two nodes can be connected by several labeled edges signifying their phonetic, collocational, syntactic, orthographic, semantic and various other kinds of similarities. Studies on such a network can reveal a holistic picture of the interaction patterns between the words, thereby providing a unified model of grammar at different levels of linguistic structure.

References

1. M. E. Adilson, A. P. S. de Moura, Y. C. Lai, and P. Dasgupta. Topology of the conceptual network of language. *Physical Review E*, 65(065102):1–4, 2002.
2. A. Agarwal, S. Chakrabarti, and S. Aggarwal. Learning to rank networked entities. In *KDD*, 2006.
3. A. Akmajian. *Linguistics. Am introduction to language and communication.* MIT Press, Cambridge, MA, 1995.
4. A. Albright and B. Hayes. Rules vs. analogy in english past tenses: A computational/experimental study. *Cognition*, 90:119–161, 2003.
5. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
6. C. Biemann. Chinese whispers - an efficient graph clustering algorithm and its application to natural language processing problems. In *Proceedings of TextGraphs: the Second Workshop on Graph Based Methods for Natural Language Processing*, pages 73–80, New York City, June 2006. Association for Computational Linguistics.
7. C. Biemann. Unsupervised part-of-speech tagging employing efficient graph clustering. In *Proceedings of the COLING/ACL 2006 Student Research Workshop*, pages 7–12, Sydney, Australia, July 2006. Association for Computational Linguistics.
8. C. Biemann, I. Matveeva, R. Mihalcea, and D. Radev, editors. *Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing.* Association for Computational Linguistics, Rochester, NY, USA, 2007.
9. S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *CNIS*, 30(1–7):107–117, 1998.
10. N. Chomsky. *The Minimalist Program.* MIT Press, Cambridge, MA, 1995.
11. M. Choudhury, M. Thomas, A. Mukherjee, A. Basu, and N. Ganguly. How difficult is it to develop a perfect spell-checker? a cross-linguistic analysis through complex network approach. In *Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing*, pages 81–88, Rochester, NY, USA, 2007. Association for Computational Linguistics.
12. A. Clark. Inducing syntactic categories by context distribution clustering. In C. Cardie, W. Daelemans, C. Nédellec, and E. T. K. Sang, editors, *Proceedings of the Fourth Conference on Computational Natural Language Learning and of the Second Learning Language in Logic Workshop, Lisbon, 2000*, pages 91–94. Association for Computational Linguistics, Somerset, New Jersey, 2000.
13. A. M. Collins and M. R. Quillian. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Memory*, 8:240–247, 1969.
14. W. Croft. *Typology and universals.* Cambridge University Press, Cambridge, MA, 1990.
15. B. de Boer. Self-organisation in vowel systems. *Journal of Phonetics*, 28(4):441–465, 2000.
16. I. S. Dhillon, S. Mallela, and D. S. Modha. Information-theoretic co-clustering. In *Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD-2003)*, pages 89–98, 2003.
17. W. B. Dolan, L. Vanderwende, and S. Richardson. Automatically deriving structured knowledge base from on-line dictionaries. In *Proceedings of the Pacific Association for Computational Linguistics*, 1993.

18. S. N. Dorogovtsev and J. F. F. Mendes. Language as an evolving word Web. *Proceedings of the Royal Society of London B*, 268(1485):2603–2606, December 22, 2001.
19. G. Erkan and D. Radev. LexRank: Graph-based lexical centrality as salience in text summarization. *JAIR*, 22:457–479, December 4, 2004.
20. C. Felbaum. *Wordnet, an Electronic Lexical Database for English.* MIT Press, Cambridge, MA, 1998.
21. R. Ferrer-i-Cancho. The structure of syntactic dependency networks: insights from recent advances in network theory. In L. V. and A. G., editors, *Problems of quantitative linguistics*, pages 60–75. 2005.
22. R. Ferrer-i-Cancho. Why do syntactic links not cross? *Europhysics Letters*, 76:1228–1235, 2006.
23. R. Ferrer-i-Cancho, A. Capocci, and G. Caldarelli. Spectral methods cluster words of the same class in a syntactic dependency network. *International Journal of Bifurcation and Chaos*, 17(7), 2007.
24. R. Ferrer-i-Cancho and R. V. Sole. The small world of human language. *Proceedings of The Royal Society of London. Series B, Biological Sciences*, 268(1482):2261–2265, November 2001.
25. R. Ferrer-i-Cancho and R. V. Sole. Two regimes in the frequency of words and the origin of complex lexicons: Zipf's law revisited. *Journal of Quantitative Linguistics*, 8:165–173, 2001.
26. R. Ferrer-i-Cancho and R. V. Sole. Patterns in syntactic dependency networks. *Physical Review E*, 69(051915), 2004.
27. S. Finch and N. Chater. Bootstrapping syntactic categories using statistical methods. In *Background and Experiments in Machine Learning of Natural Language: Proceedings of the 1st SHOE Workshop*, pages 229–235. Katholieke Universiteit, Brabant, Holland, 1992.
28. D. Freitag. Toward unsupervised whole-corpus tagging. In *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, page 357, Morristown, NJ, USA, 2004. Association for Computational Linguistics.
29. M. Galley and K. McKeown. Improving word sense disambiguation in lexical chaining. In *IJCAI*, 2003.
30. M. Gamon. Graph-based text representation for novelty detection. In *Proceedings of the Workshop on TextGraphs at HLT-NAACL*, pages 17–24, 2006.
31. S. Gauch and R. Futrelle. Experiments in Automatic Word Class and Word Sense Identification for Information Retrieval. In *Proceedings of the 3rd Annual Symposium on Document Analysis and Information Retrieval*, pages 425–434, Las Vegas, NV, April 1994.
32. M. Gell-Mann. Language and complexity. In J. W. Minett and W. S.-Y. Wang, editors, *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics*. City University of Hong Kong Press, July 2005.
33. D. Gibson, J. M. Kleinberg, and P. Raghavan. Inferring Web communities from link topology. In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia*, pages 225–234, 1998.
34. A. B. Goldberg and J. Zhu. Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. In *HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing*, 2006.
35. J. H. Greenberg and J. J. Jenkins. Studies in the psychological correlates of the sound system of american english. *Word*, 20:157–177, 1964.

36. T. Gruenenfelder and D. B. Pisoni. *Modeling the mental lexicon as a complex graph.* Ms. Indiana University, 2005.
37. Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating Web spam with TrustRank. In *VLDB*, pages 576–587, 2004.
38. A. D. Haghighi, A. Y. Ng, and C. D. Manning. Robust textual inference via graph matching. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 387–394, Morristown, NJ, USA, 2005. Association for Computational Linguistics.
39. Z. S. Harris. *Mathematical Structures of Language.* Wiley, 1968.
40. M. D. Hauser, N. Chomsky, and W. T. Fitch. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298:1569–1579, 11 2002.
41. R. F. i Cancho, A. Mehler, O. Pustylnikov, and A. Díaz-Guilera. Correlations in the organization of large-scale syntactic dependency networks. In *TextGraphs-2: Graph-Based Algorithms for Natural Language Processing*, pages 65–72. Association for Computational Linguistics, 2007.
42. Y. Itoh and S. Ueda. The ising model for changes in word ordering rules in natural languages. *Physica D: Nonlinear Phenomena*, 198(3-4):333–339, 11 2004.
43. J. Jannink and G. Wiederhold. Thesaurus entry extraction from an on-line dictionary. In *Fusion*, 1999.
44. B. Jedynak and D. Karakos. Unigram language models using diffusion smoothing over graphs. In *Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing*, pages 33–36, Rochester, NY, USA, 2007. Association for Computational Linguistics.
45. V. Kapatsinski. Sound similarity relations in the mental lexicon: Modeling the lexicon as a complex network. *Speech Research Lab Progress Report, Indiana University*, 2006.
46. V. Kapustin and A. Jamsen. Vertex degree distribution for the graph of word co-occurrences in Russian. In *Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing*, pages 89–92, Rochester, NY, USA, 2007. Association for Computational Linguistics.
47. J. Ke, M. Ogura, and W. S.-Y. Wang. Optimization models of sound systems using genetic algorithms. *Computational Linguistics*, 29(1):1–18, 2003.
48. J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of ACM*, 46, 1999.
49. R. Kumar, J. Novak, P. Raghavan, and A. Tomkins. Structure and evolution of blogspace. *Communications of the ACM*, 47(12):35–39, 2004.
50. M. Lesk. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *SIGDOC*, 1986.
51. J. Liljencrants and B. Lindblom. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language*, 48:839–862, 1972.
52. H. Liljenstrom. *Micro Meso Macro: Addressing Complex Systems Couplings*. World Scientific Publishing Co. Pvt. Ltd., 2005.
53. P. A. Luce and D. B. Pisoni. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19:1–36, 1998.
54. I. Maddieson. *Patterns of sounds.* Cambridge University Press, Cambridge, 1984.
55. W. Marslen-Wilson. Activation, competition, and frequency in lexical access. In: G.T.M. Altmann (ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives,* Cambridge, MA: MIT, pages 148–173, 1990.

56. R. McDonald, F. Pereira, K. Ribarov, and J. Hajič. Non-projective dependency parsing using spanning tree algorithms. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 523–530, Morristown, NJ, USA, 2005. Association for Computational Linguistics.
57. R. Mihalcea. Graph-based ranking algorithms for large vocabulary word sense disambiguation. In *HTL-EMNLP*, 2005.
58. R. Mihalcea and D. Radev. Graph-based algorithms for information retrieval and natural language processing. Tutorial at HLT/NAACL 2006, 2006.
59. R. Mihalcea and D. Radev, editors. *Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing*. Association for Computational Linguistics, 2006.
60. R. Mihalcea and P. Tarau. TextRank: Bringing order into texts. In *EMNLP 2004*, 2004.
61. R. Mihalcea, P. Tarau, and E. Figa. Pagerank on semantic networks with applications to word sense disambiguation. In *COLING*, 2004.
62. G. A. Miller and W. G. Charles. Contextual Correlates of Semantic Similarity. *Language and Cognitive Processes*, 6(1):1–28, 1991.
63. G. A. Miller and P. M. Gildea. How children learn words. *Scientific American*, 257(3):86–91, 1987.
64. A. Mukherjee, M. Choudhury, A. Basu, and N. Ganguly. Modeling the co-occurrence principles of the consonant inventories: A complex network approach. *International Journal of Modern Physics C*, 18(2):281–295, 2007.
65. A. Mukherjee, M. Choudhury, A. Basu, and N. Ganguly. Self-organization of sound inventories: Analysis and synthesis of the occurrence and co-occurrence networks of consonants. *Journal of Quantitative Linguistics*, to appear.
66. J. Nath, M. Choudhury, A. Mukherjee, C. Biemann, and N. Ganguly. Unsupervised parts-of-speech induction for Bengali. In *Proceedings of the Sixth International Language Resources and Evaluation Conference (LREC)*, 2008.
67. D. Nettle. Using social impact theory to simulate language change. *Lingua*, 108:95–117, 1999.
68. H. G. Nusbaum, D. B. Pisoni, and C. K. Davis. Sizing up the hoosier mental lexicon: Measuring the familiarity of 20,000 words, indiana university. *Research on Speech Perception Progress Report No. 10*, pages 357–376, 1984.
69. B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 271–278, Barcelona, Spain, July 2004.
70. S. Pinker. *The Language Instinct: How the Mind Creates Language*. HarperCollins, New York, 1994.
71. S. Pinker and A. Price. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28:195–247, 1988.
72. R. Rapp. A practical solution to the problem of automatic part-of-speech induction from text. In *Conference Companion Volume of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL-05)*, Ann Arbor, Michigan, USA, 2005.
73. M. Richardson, A. Prakash, and E. Brill. Beyond PageRank: Machine learning for static ranking. In *WWW*, pages 707–715, 2006.

74. H. Schtze. Part-of-speech induction from scratch. In *Proceedings of the 31st annual meeting on Association for Computational Linguistics*, pages 251–258, Morristown, NJ, USA, 1993. Association for Computational Linguistics.
75. H. Schtze. Distributional part-of-speech tagging. In *Proceedings of the 7th Conference on European chapter of the Association for Computational Linguistics*, pages 141–148, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.
76. J.-L. Schwartz, L.-J. Boë, N. Vallée, and C. Abry. The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, 25:255–286, 1997.
77. M. Sigman and G. A. Cecchi. Global organization of the wordnet lexicon. *Proceedings of the National Academy of Science*, 99(3):1742–1747, 2002.
78. M. M. Soares, G. Corso, and L. S. Lucena. The network of syllables in portuguese. *Physica A: Statistical Mechanics and its Applications*, 355(2-4):678–684, 2005.
79. Z. Solan, D. Horn, E. Ruppin, and S. Edelman. Unsupervised learning of natural languages. *Proceedings of National Academy of Sciences*, 102(33):11629–11634, 2005.
80. L. Steels. Language as a complex adaptive system. In *Proceedings of PPSN VI*, pages 17–26, 2000.
81. D. Steriade. Knowledge of similarity and narrow lexical override. *BLS*, 29:583–598, 2004.
82. M. Steyvers and J. B. Tenenbaum. The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78, 2005.
83. M. Tamariz. *Exploring the Adaptive Structure of the Mental Lexicon*. PhD thesis, Department of theoretical and applied linguistics, Univerisity of Edinburgh, 2005.
84. K. Toutanova, C. D. Manning, and A. Y. Ng. Learning random walk models for inducing word dependency distributions. In *ICML '04: Proceedings of the twenty-first international conference on Machine learning*, page 103, New York, NY, USA, 2004.
85. J. Véronis. HyperLex: Lexical cartography for information retrieval. *Computer Speech and Language*, 18(3):223–252, 2004.
86. M. S. Vitevitch. *Phonological Neighbors in a Small World*. Ms. University of Kansas, 2004.
87. D. Widdows and B. Dorow. A graph model for unsupervised lexical acquisition. In *Proceedings of COLING*, 2002.