The Topology of Synonymy and Homonymy Networks

James Gorman and James R. Curran
School of Information Technologies
University of Sydney
NSW 2006, Australia
{jgorman2,james}@it.usyd.edu.au

Abstract

Semantic networks have been used successfully to explain access to the mental lexicon. Topological analyses of these networks have focused on acquisition and generation. We extend this work to look at models that distinguish semantic relations. We find the scale-free properties of association networks are not found in synonymy-homonymy networks, and that this is consistent with studies of childhood acquisition of these relationships. We further find that distributional models of language acquisition display similar topological properties to these networks.

1 Introduction

Semantic networks have played an important role in the modelling of the organisation of lexical knowledge. In these networks, words are connected by graph edges based on their semantic relations. In recent years, researchers have found that many semantic networks are *small-world*, *scale-free* networks, having a high degree of structure and a short distance between nodes (Steyvers and Tenenbaum, 2005).

Early models were taxonomic and explained some aspects of human reasoning (Collins and Quillian, 1969) (and are still used in artificial reasoning systems), but were replaced by models that focused on general graph structures (e.g. Collins and Loftus, 1975). These better modelled many observed phenomena but explained only the searching of semantic space, not its generation or properties that exist at a whole-network level.

Topological analyses, looking at the statistical regularities of whole semantic networks, can be used to model phenomena not easily explained from the smaller scale data found in human experiments. These networks are typically formed from corpora, expert compiled lexical resources, or human word-association data.

Existing work has focused language acquisition (Steyvers and Tenenbaum, 2005) and generation (Cancho and Solé, 2001). These models use the general notion of semantic *association* which subsumes all specific semantic relations, e.g. synonymy.

There is evidence that there are distinct cognitive processes for different semantic relations (e.g. Casenhiser, 2005). We perform a graph analysis of *synonymy*, nearness of meaning, and *homonymy*, shared lexicalisation.

We find that synonymy and homonymy produce graphs that are topologically distinct from those produced using association. They still produce small-world networks with short path lengths but lack scale-free properties. Adding edges of different semantic relations, in particular hyponymy, produces graphs more similar to the association networks. We argue our analyses consistent with other semantic network models where nodes of a common type share edges of different types (e.g. Collins and Loftus, 1975).

We further analyse the distributional model of language acquisition. We find that it does not well explain whole-language acquisition, but provides a model for synonym and homonym acquisition.

Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition, pages 73–80,
Prague, Czech Republic, June 2007

2 Graph Theory

Our overview of graph theory follows Watts (1999). A graph consists of a set of n *vertices* (*nodes*) and a set of *edges*, or *arcs*, which join pairs of vertices. Edges are *undirected* and arcs are *directed*. Edges and arcs can be *weighted* or *unweighted*, with weights indicating the relative strength or importance of the edges. We will only consider unweighted, undirected networks. Although there is evidence that semantic relations are both directed (Tversky, 1977) and weighted (Collins and Loftus, 1975), we do not have access to this information in a consistent and meaningful format for all our resources.

Two vertices connected by an edge are called *neighbours*. The *degree* k of a vertex is the count of it neighbours. From this we measure the average degree $\langle k \rangle$ for the graph and the degree distribution $P(k)$ for all values of k. The degree distribution is the probability of a vertex having a degree k.

The *neighbourhood* Γ_v of a vertex v is the set of all neighbours of v not including v. The neighbourhood Γ_S of a subgraph S is the set of all neighbours of S, not including the members of S.

The *distance* between any two vertices is the *shortest path length*, or the minimum number of edges that must be traversed, to reach the first from the second. The *characteristic path length* L is the average distance between vertices.[1] The *diameter* D of a graph is the maximum shortest path length between any two vertices. At most D steps are required to reach any vertex from any other vertex but, on average, only L are required.

For very large graphs, calculating the values for L and D is computationally difficult. We instead sample $n' \ll n$ nodes and find the mean values of L and D across the samples. The diameter produced will always be less than or equal to the true diameter. We found $n' = 100$ to be most efficient.

It is not a requirement that every vertex be reachable from every other vertex and in these cases both L and D will be infinite. In these cases we analyse the largest connected subgraph.

[1] Here we follow Steyvers and Tenenbaum (2005) as it is more commonly used in the cognitive science literature. Watts (1999) defines the characteristic path length as the *median* of the means of shortest path lengths for each vertex.

2.1 Small-world Networks

Traditional network models assume that networks are either completely random or completely regular. Many natural networks are somewhere between these two extremes. These *small-world* networks a have the high degree of clustering of a regular *lattice* and the short average path length of a random network (Watts and Strogatz, 1998). The clustering is indicative of organisation, and the short paths make for easier navigation.

The *clustering coefficient* C_v is used to measure the degree of clustering around a vertex v:

$$C_v = \frac{|E(\Gamma_v)|}{\binom{k_v}{2}}$$

where $|E(\Gamma_v)|$ is the number of edges in the neighbourhood Γ_v and $\binom{k_v}{2}$ is the total number of possible edges in Γ_v. The clustering coefficient C of a graph is the average over the coefficients of all the vertices.

2.2 The Scale of Networks

Amaral et al. (2000) describe three classes of small world networks based on their degree distributions:

Scale-free networks are characterised by their degree distribution decaying as a power law, having a small number of vertices with many links (*hubs*) and many vertices with few links. Networks in this class include the internet (Faloutsos et al., 1999) and semantic networks (Steyvers and Tenenbaum, 2005).

Broad-scale networks are characterised by their degree distribution decaying as a power law followed by a sharp cut-off. This class includes collaborative networks (Watts and Strogatz, 1998).

Single-scale networks are characterised by fast decaying degree distribution, such exponential or Gaussian, in which hubs are scarce or nonexistent. This class includes power grids (Watts and Strogatz, 1998) and airport traffic (Amaral et al., 2000).

Amaral et al. (2000) model these differences using a constrained preferential attachment model, where new nodes prefer to attach to highly connected nodes. Scale-free networks result when there are no constraints. Broad-scale networks are produced when ageing and cost-to-add-link constraints are added, making it more difficult to produce very high degree hubs. Single-scale networks occur when these constraints are strengthened. This is one of several models for scale-free network generation, and different models will result in different internal structures and properties (Keller, 2005).

3 Semantics Networks

Semantic networks represent the structure of human knowledge through the connections of words. Collins and Quillian (1969) proposed a taxonomic representation of knowledge, where words are connected by hyponym relations, like in the WordNet noun hierarchy (Fellbaum, 1998). While this structure predicted human reaction times for verifying facts it allows only a limited portion of knowledge to be expressed. Later models represented knowledge as semi-structured networks, and focused on explaining performance in memory retrieval tasks. One such model is *spreading-activation*, in which the degree to which a concept is able to be recalled is related to its similarity both to other concepts in general and to some particular *prime* or primes (Collins and Loftus, 1975). In this way, if one is asked to name a red vehicle, fire truck is more likely response than car: while both are strongly associated with vehicle, fire truck is more strongly associated with red than is car.

More recently, graph theoretic approaches have examined the topologies of various semantic networks. Cancho and Solé (2001) examine graphs of English modelled from the British National Corpus. Since co-occurrence is non-trivial — words in a sentence must share some semantic content for the sentence to be coherent — edges were formed between adjacent words, with punctuation skipped. Two graphs were formed: one from all co-occurrences and the other from only those co-occurrences with a frequency greater than chance. Both models produced scale-free networks. They find this model compelling for word choice during speech, noting function words are the most highly connected. These give structure without conveying significant meaning, so can be omitted without rendering a sentence incoherent, but when unavailable render speech non-fluent. This is consistent with work by Albert et al. (2000) showing that scale-free networks are tolerant to random deletion but sensitive to targeted removal of highly connected vertices.

Sigman and Cecchi (2002) investigate the structure of WordNet to study the effects of nounal polysemy on graph navigation. Beginning with synsets and the hyponym tree, they find adding polysemy both reduces the characteristic path length and increases the clustering coefficient, producing a small-world network. They propose, citing word priming experiments as evidence, that these changes in structure give polysemy a role in metaphoric thinking and generalisation by increasing the navigability of semantic networks.

Steyvers and Tenenbaum (2005) examine the growth of semantic networks using graphs formed from several resources: the free association index collected by Nelson et al. (1998), Wordnet and the 1911 Roget's thesaurus. All these produced scale-free networks, and, using an age of acquisition and frequency weighted preferential attachement model, show that this corresponds to age-of-acquisition norms for a small set of words. This is compared to networks produced by Latent Semantic Analysis (LSA, Landauer and Dumais, 1997), and conclude that LSA is an inadequate model for language growth as it does not produce the same scale-free networks as their association models.

3.1 Synonymy and Homonymy

While there have been many studies using human subjects on the acquisition of particular semantic relations, there have been no topological studies differentiating these from the general notion of semantic *association*. This is interesting as psycholinguistic studies have shown that semantic relationships are distinguishable (e.g. Casenhiser, 2005). Here we consider *synonymy* and *homonymy*.

There are very few cases of true synonymy, where two words are substitutable in all contexts. Near-synonymy, where two words share some close common meaning, is more common. Sets of synonyms can be grouped together into *synsets*, representing a common idea.

Homonymy occurs when a word has multiple meanings. Formally, homonymy is occurs when words do not share an etymological root (in linguistics) or when the distinction between meanings is coarse (in cognitive science). When the words share a root or meanings are close, the relationship is called *polysemy*. This distinction is significant in language acquisition, but as yet little research has been performed on the learning of polysemes (Casenhiser, 2005). It is also significant for Natural Language Processing. The effect of disambiguating homonyms is markedly different from polysemes in Information Retrieval (Stokoe, 2005).

We do not have access to these distinctions, as they are not available in most resources, nor are there techniques to automatically acquire these distinctions (Kilgarriff and Yallop, 2000). For simplicity, will conflate the categories under homonymy.

There have been several studies into synonymy and homonymy acquisition in children, and these have shown that it lags behind vocabulary growth (Doherty and Perner, 1998; Garnham et al., 2000). A child will associate both rabbit and bunny with the same concept, but before the age of four, most children have difficulty in choosing the word bunny if they have already been presented with the word rabbit. Similarly, a young child asked to point to two pictures that have the same name but mean different things will have difficulty, despite knowing each of the things independently.

Despite this improvement with age, there are tendencies for language to avoid synonyms and homonyms as a more general principle of economy (Casenhiser, 2005). This is balanced by the utility of ambiguous relations for mental navigation (Sigman and Cecchi, 2002) which goes some way to explaining why they still play such a large role in language.

4 The Topology of Synonymy and Homonymy Relations

For each of our resources we form a graph based on the relations between lexical items. This differs to the earlier work of Sigman and Cecchi (2002), who use synsets as vertices, and Steyvers and Tenenbaum (2005) who use both lexical items and synsets..

This is motivated largely by our automatic acquisition techniques, and also by human studies, in which we can only directly access relationships between words. This also allows us to directly compare resources where we have information about synsets to those without. We distinguish parts of speech as disambiguation across them is relatively easy psychologically (Casenhiser, 2005) and computationally (e.g. Ratnaparkhi, 1996).

4.1 Lexical Semantic Resources

A typical resource for providing this information are manually constructed lexical semantic resources. We will consider three: Roget's, WordNet and Moby

Roget's thesaurus is a common language thesaurus providing a hierarchy of synsets. Synsets with the same general or overlapping meaning and part of speech are collected into paragraphs. The parts of speech covered are nouns, verbs, adjectives, adverbs, prepositions, phrases, pronouns, interjections, conjunctions, and interrogatives. Paragraphs with similar meaning are collated by part of speech into labeled categories. Categories are then collected into classes using a three-tiered hierarchy, with the most general concept at the top. Where a word has several senses, it will appear in several synsets. Several editions of Roget's have been released representing the change in language since the first edition in 1852. The last freely available edition is the 1911, which uses outdated vocabulary, but the global topology has not changed with more recent editions (Old, 2003). As our analysis is not concerned with the specifics of the vocabulary, this is the edition we will use. It consists of a vocabulary of 29,460 nouns, 15,173 verbs, 13,052 adjectives and 3,005 adverbs.

WordNet (Fellbaum, 1998) is an electronic lexical database. Like Roget's, it main unit of organisation is the synset, and a word with several senses will appear in several synsets. These are divided into four parts of speech: nouns, verbs, adjectives and adverbs. Synsets are connected by semantic relationships, e.g antonymy, hyponymy and meronym. WordNet 2.1 provides a vocabulary of 117,097 nouns, 11,488 verbs, 22,141 adjectives and 4,601 adverbs.

The Moby thesaurus provides synonymy lists for over 30,000 words, with a total vocabulary of 322,263 words. These lists are not distinguished by part of speech. A separate file is supplied containing part of speech mappings for words in the vocabulary. We extracted separate synonym lists for nouns, verbs, adjectives and adverbs using this list combined with WordNet part of speech information.[2] This produces a vocabulary of 42,821 nouns, 11,957 verbs, 16,825 adjectives and 3,572 adverbs.

Table 1 presents the statistics for the largest con-

[2]http://aspell.sourceforge.net/wl/

	Roget's				WordNet				Moby			
	Noun	Verb	Adj	Adv	Noun	Verb	Adj	Adv	Noun	Verb	Adj	Adv
n	15,517	8,060	6,327	626	11,746	6,506	4,786	62	42,819	11,934	16,784	3501
$\langle k \rangle$	8.97	8.46	7.40	7.17	4.58	6.34	5.16	4.97	34.65	51.98	39.26	16.07
L	6.5	6.0	6.4	10.5	9.8	6.0	9.5	5.6	3.7	3.1	3.4	3.7
D	21.4	17	17	31	27	15.3	26.4	14	9.6	9.8	9.3	9.8
C	0.74	0.68	0.69	0.77	0.63	0.62	0.66	0.57	0.60	0.49	0.57	0.55
L_r	4.7	4.5	4.6	3.5	6.3	5.0	5.9	3.3	3.4	2.8	2.9	3.2
D_r	8.5	8.4	9.0	7	13.3	10.1	11.8	8	5.5	5	5	6
C_r	0.00051	0.0011	0.0012	0.0090	0.00036	0.00099	0.00094	0.028	0.00081	0.0043	0.0023	0.0047

Table 1: Topological statistics for nouns, verbs, adjectives and adverbs for our three gold standard resources

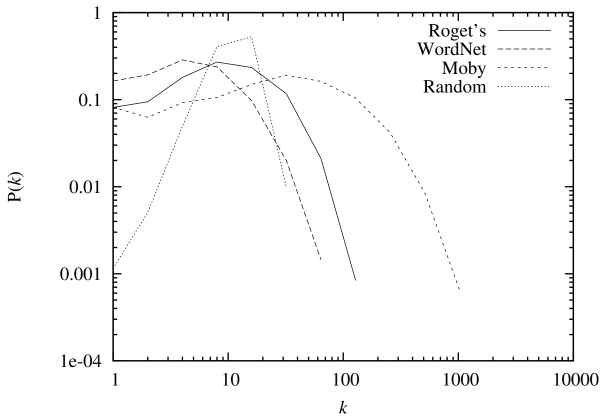

Figure 1: Degree distributions for nouns

nected subgraph for the four parts of speech considered, along with statistics for random graphs of equivalent size and average degree (subscript r). In all cases the clustering coefficient is significantly higher than that for the random graph. While the characteristic path length and diameter are larger than for the random graphs, they are still short in comparison to an equivalent latice. This, combined with the high clustering coefficient, indicates that they are producing small-world networks. The diameter is larger still than for the random graphs. Together these indicate a more lattice like structure, which is consistent with the intuition that dissimilar words are unlikely to share similar words. This is independent of part of speech.

Figure 1 shows the degree distributions for nouns, and for a random graph plotted on log-log axes. Other parts of speech produce equivalent graphs. These clearly show that we have not produced scale-free networks as we are not seeing straight line power law distributions. Instead we are seeing what is closer to single- or broad-scale distributions.

The differences in the graphs is explained by the

	WordNet		Roget's		
		Hyp	Synset	Para	Cat
n	11,746	118,264	15,517	27,989	29,431
$\langle k \rangle$	4.58	6.61	8.97	26.84	140.36
L	9.8	6.3	6.5	4.3	2.9
D	27	16.4	21.4	12.6	7
C	0.63	0.74	0.74	0.85	0.86

Table 2: Effect of adding hyponym relations

granularity of the synonymy relations presented, as indicated by the characteristic path length. WordNet has fine grained synsets and the smallest characteristic path length, while Moby has coarse grained synonyms and the largest characteristic path length.

4.2 Synonymy-Like Relations

Having seen that synonymy and homonymy alone do not produce scale-free networks, we investigate the synonymy-like relations of *hyponymy* and topic relatedness. Hyponymy is the IS-A class subsumption relationship and occurs between noun synsets in WordNet. Topic relatedness occurs in the grouping of synsets in Roget's in paragraphs and categories.

Table 2 compares adding hyponym edges to the graph of WordNet nouns and increasing the granularity of Roget's synsets using edges between all words in a paragraph or category. Adding hyponymy relations increases the connectivity of the graph significantly and there are no longer any disconnected subgraphs. At the same time the diameter is nearly halved and characteristic path length reduce one third, but average degree only increases by one third. To achieving the same reduction in path length and diameter by the granularity of Roget's requires the average degree to increase by nearly three times.

Figure 2 shows the degree distributions when hyponyms are added to WordNet nouns and the granularity of Roget's is increased. Roget's category level graph is omitted for clarity. We can see that the orig-

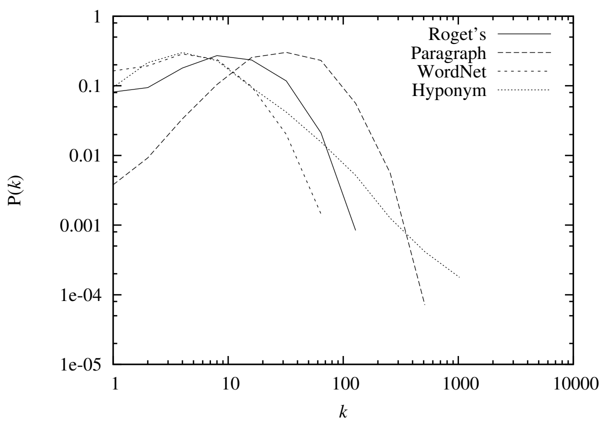

Figure 2: Degree distributions adding hyponym relations to nouns

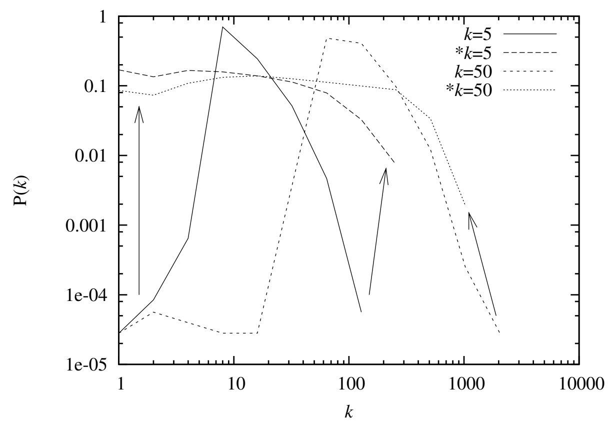

Figure 3: Degree distributions of Jaccard

inally broad-scale structure of the Roget's distribution is tending to have a more gaussian distribution. The addition of hyponyms produces a power law distribution for $k > 10$ of $P(k) \approx k^{-1.7}$.

Additional constraints on attachment reduce the ability of networks to be scale-free (Amaral et al., 2000). The difference between synonymy-homonymy networks and association networks can be explained by this. Steyvers and Tenenbaum (2005) propose a plausible attachment model for their association networks which has no additional constraint function. If we use the tendency for languages to avoid lexical ambiguity from synonymy and homonymy as a constraint to the production of edges we will produce broad-scale networks rather than scale-free networks.

As hyponymy is primarily semantic and does not produce lexical ambiguity, adding hyponym edges weakens the constraint on ambiguity, producing a scale-free network. Generalising synonymy to include topicality weakens the constraints, but at the same time reduces preference in attachment. The results of this is the gaussian-like distribution with very few low degree nodes. The difference between this thesaurus based topicality and that found in human association data is that human association data only includes the most similar words.

5 Distributional Similarity Networks

Lexical semantic resources can be automatically extracted using distributional similarity. Here words are projected into a vector space using the contexts in which they appear as axes. Contexts can be as wide as document (Landauer and Dumais, 1997) or close as grammatical dependencies (Grefenstette, 1994). The distance between words in this space approximates the similarity measured by synonymy.

We use the noun similarities produced by Gorman and Curran (2006) using the *weighted Jaccard* measure and the *t-test* weight and grammatical relations extracted from their LARGE corpus, the method found to perform best against their gold-standard evaluation. Only words with a corpus frequency higher than 100 are included. This method is comparable to that used in LSA, although using grammatical relations as context produces similarity much more like synonymy than those taken at a document level (Kilgarriff and Yallop, 2000).

Distributional similarity produces a list of vocabulary words, their similar neighbours and the similarity to the neighbours. These lists approximate synonymy by measuring substitutability in context, and do not only find synonyms as near neighbours as both antonyms and hyponyms are frequently substitutable in a grammatical context (Weeds, 2003). From this we generate graphs by taking either the k nearest neighbours to each word (k-NN), or by using a threshold. To produce a threshold we take the mean similarity of the k^{th} neighbour of all words (*k-NN). We compare both these methods.

Figure 3 compares the degree distributions of these. Using k-NN produces a degree distribution that is close to a Gaussian, where as *k-NN produces a distribution much more like that of our expert compiled resources. This is unsurprising when the distribution of distributional distances is considered. Some words will have many near neighbours,

	Roget's	WordNet	Hyp	k-NN	*k-NN
n	15,517	11,746	118,264	35,592	19,642
$\langle k \rangle$	8.97	4.58	6.61	8.26	13.86
L	6.5	9.8	6.3	6.2	6.4
D	21.4	27	16.4	12	25.6
C	0.74	0.63	0.74	0.18	0.37

Table 3: Comparing nouns in expert and distributional resources

and other few. In the first case, k-NN will fail to include some near neighbours, and in the second will include some distant neighbours that are note semantically related. This result is consistent between $k = 5$ and 50. Introduction of random edges from the noise of distant neighbours reduces the diameter and missing near neighbours reduces the clustering coefficient (Table 3).

In Table 3 we also compare these to noun synonymy in Roget's, and to synonymy and hyponymy in WordNet. Distributional similarity (*k-NN) produces a network with similar degree, characteristic path length and diameter. The clustering coefficient is much less than that from expert resources, is still several orders of magnitude larger than an equivalent random graph (Table 1).

Figure 4 compares a distributional network to networks WordNet and Moby. We can see the same broad-scale in the distributional and synonym networks, and a distinct difference with the scale-free WordNet hyponym distribution.

The distributional similarity distribution is similar to that found in networks formed from LSA by Steyvers and Tenenbaum (2005). Steyvers and Tenenbaum hypothesise that the distributions produced by LSA might be due more to frequency distribution effects that correct language modelling.

In light of our analysis of synonymy relations, we propose a new explanation. Given that: distributional similarity has been shown to approximate the semantic similarity in synonymy relations found in thesaurus type resources (Curran, 2004); distributional similarity produces networks with similar statistical properties to those formed by synonym and homonymy relations; and, the synonym and homonymy relations found in thesauri produce networks with different statistical properties to those found in the association networks analysed by Steyvers and Tenenbaum; it can be plausibly hypothesised that distributional techniques are modeling the acquisition of synonyms and homonyms, rather than all semantic relationships.

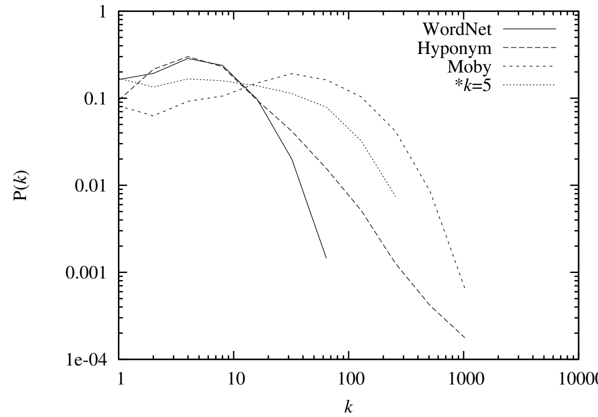

Figure 4: Degree distributions for nouns

This is given further credence by experimental findings that acquisition of homonyms occurs at a different rate to the acquisition of vocabulary. This indicates that there are different mechanisms for learning the meaning of lexical items and learning to relate the meanings of lexical items. Any whole-language model would then be composed of a common set of lexical items related by disparate relations, such as synonymy, homonymy and hyponymy. This type of model is predicted by spreading activation (Collins and Loftus, 1975).

It is unfortunate that there is a lack of data with which to validate this model, or our constraint model, empirically. This should not prevent further analysis of network models that distiguish semantic relations, so long as this limitation is understood.

6 Conclusion

Semantic networks have been used successfully to explain access to the mental lexicon. We use both expert-compiled and automatically extracted semantic resources, we compare the networks formed from semantic association and synonymy and homonymy. These relations produce small-world networks, but do not share the same scale-free properties as for semantic association.

We find that this difference can be explained using a constrained attachment model informed by childhood language acquisition experiments. It is also predicted by spreading-activation theories of semantic access where a common set of lexical items is connected by a disparate set of relations. We further find that distributional models of language acquisition produce relations that approximate synonymy and networks topologically similar to synonymy-homonymy networks.

7 Acknowledgements

We would like to thank the anonymous reviewers for their helpful feedback and corrections. This work has been supported by the Australian Research Council under Discovery Projects DP0453131 and DP0665973.

References

Réka Albert, Hawoong Jeong, and Albert-László Barabási. 2000. Error and attack tolerance of complex networks. *Nature*, 406:378–381.

Luís A. Nunes Amaral, Antonio Scala, Marc Barthélémy, and H. Eugene Stanley. 2000. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21):11149–11152, October 10.

Ramon F. i Cancho and Ricard V. Solé. 2001. The small world of human language. *Proceedings of The Royal Society of London. Series B, Biological Sciences*, 268(1482):2261–2265, November.

Devin M. Casenhiser. 2005. Children's resistance to homonymy: an experimental study of pseudohomonyms. *Journal of Child Language*, 32:319–343.

Allan M. Collins and Elizabeth F. Loftus. 1975. A spreading-activation theory of semantic processing. *Psychological review*, 82(6):407–428.

Allan M. Collins and M. Ross Quillian. 1969. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8:240–247.

James R. Curran. 2004. *From Distributional to Semantic Similarity*. Ph.D. thesis, University of Edinburgh.

Martin Doherty and Josef Perner. 1998. Metalinguistic awareness and theory of mind: just two words for the same thing? *Congitive Development*, 13:279–305.

Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. 1999. On power-law relationships of the internet topology. In *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 251–262.

Christiane Fellbaum, editor. 1998. *WordNet: an electronic lexical database*. The MIT Press, Cambridge, MA, USA.

Wendy A. Garnham, Julie Brooks, Alan Garnham, and Anne-Marie Ostenfeld. 2000. From synonyms to homonyms: exploring the role of metarepresentation in language understanding. *Developmental Science*, 3(4):428–441.

James Gorman and James R. Curran. 2006. Scaling distributional similarity to large corpora. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia, 17–21 July.

Gregory Grefenstette. 1994. *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers, Boston.

Evelyn F. Keller. 2005. Revisiting "scale-free" networks. *Bioessays*, 27(10):1060–1068, October.

Adam Kilgarriff and Colin Yallop. 2000. What's in a thesaurus? In *Proceedings of the Second International Conference on Language Resources and Evaluation*, pages 1371–1379.

Thomas K. Landauer and Susan T. Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240, April.

Douglas L. Nelson, Cathy L. McEvoy, and Thomas A. Schreiber. 1998. The university of south florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

L. John Old. 2003. *The Semantic Structure of Roget's, a Whole-Language Thesaurus*. Ph.D. thesis, Indiana University.

Adwait Ratnaparkhi. 1996. A maximum entropy part-of-speech tagger. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 133–142, 17–18 May.

Mariano Sigman and Guillermo A. Cecchi. 2002. Global organization of the WordNet lexicon. *Proceedings of the National Academy of Sciences*, 99(3):1742–1747.

Mark Steyvers and Joshua B. Tenenbaum. 2005. The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78.

Christopher Stokoe. 2005. Differentiating homonymy and polysemy in information retrieval. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 403–410.

Amos Tversky. 1977. Features of similarity. *Psychological Review*, 84(4):327–352, July.

Duncan J. Watts and Steven H. Strogatz. 1998. *Collective dynamics of small-world networks*, 393:440–442, 4 June.

Duncan J. Watts. 1999. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, Princeton, NJ, USA.

Julie E. Weeds. 2003. *Measures and Applications of Lexical Distributional Similarity*. Ph.D. thesis, University of Sussex.