

CS790R

Paper Review Presentation Notes

Kara DeSouza

For presentation date of 03/27/2006

Paper presented:

Motter, A. E., de Moura, A. P. S., Lai, Y.-C. and Dasgupta, P. (2002) Topology of the conceptual network of language. [*Physical Review E*](#), **65**: 065102.

Main points of the paper:

- The English language (among other languages) can be selectively modeled as a small-world network.
- Accomplished this by using the inherent semantic connections of an online thesaurus,
- Words were considered to be “nodes” and word meaning comprised the edges
- Synonyms, and only those which were both entries in their own right, and words listed under other entries, were used
- The resulting network displayed classical small-world patterning; i.e. there were a high number of clusters of similarly related words, and the path length between any two words, no matter how seemingly disparate they could be, was short (an average of about 3 links).

Methodological and Theoretical Questions discussed in seminar:

- A thesaurus may be a useful, if “loaded” tool to use when attempting to model language on a small world network, due to the inherent nature of word-connectivity of a thesaurus

-Is the resulting small-world network reasonably constructed even though antonyms and “dead end” words were presumably cut out of the selection of words analyzed to create the model?

-How does this model compare to the Ferrer-i-Cancho paper in terms of the body of language used to construct the respective models? (i.e. does using a “natural” language source such as sentences belie a different or more realistic network than using the pre-arranged words of a thesaurus?)

-Network models are currently a very “hot topic” model to impose on any number of situations, and for many real-world phenomenon, the use of these models has provided new insight into old problems. However, can and should they be applied to just anything?

-Just because our written language examples can be broken down into these models, does that mean that our experiential language is actually brought forth from our brains in the same way?

-Are the relationships between words, particularly those given as examples in the article, really as equally distributed as the article would imply? That is, is it equally as likely as that the word “nature” should evoke the word “character” as it does “universe?”

-The model is not weighted, nor does it seem to compensate for the fact that language does not develop randomly, but that from the moment of it’s creation, a new word is already linked to nodes of similar meaning, and does not have the opportunity to randomly connect to “just any node” out of the system.

-**What significance should we put on random connectivity to nodes anyway, in models of real-life systems? (Since most RL systems have inherent weights that shape the creation and maintenance of the system anyway?)

-**Although the Ferrer-i-Cancho direction of research (i.e. using “whole” language examples like sentences to create the model, rather than single words) has probably a more realistic

methodological basis for studying the network characteristics of language, if the Motter study were to be replicated, what might be learned from combining the thesaurus data with “word norm” data commonly used by linguists? (This would have the effect of pre-weighting some of the nodes by their “commonality of use in everyday language” before they ever came in contact with each other through the thesaurus links. It would not, probably, solve the “dead-ends and antonyms” issue.)

**These are issues which we only briefly discussed in class, or which after class it occurred to me could have been expounded upon with some further questioning.