Analysis of Language Resources

Fourth Lecture Hiroyuki Akama

Word Association for the Amodal Theory...(Review)

- The possibilities of defining the semantics can be founded only upon the "association" of words.
- The meaning in language can be considered as being generated from the space between terms.
- The meaning is by itself a context, a chain of words or a block of adjacent terms found in a document.
- Each word in a document can be considered as a variable or an observation instance. →The meaning will be found in a *component* that can be calculated for lexical co-occurrence data by using the probabilistic theory.

Calculating Word Association

- The coherence of the paragraphs is determined by calculating the similarity measure of the contained sentences.
 - I. Marissa forgot to bring her pillow on her camping trip. As her substitute for her pillow, she filled up her own sweater with leaves. (LSA value: 0.58) SIMILAR PARAGRAPH!?
 - II. Marissa forgot to bring her pillow on her camping trip. As her substitute for her pillow, she filled up her own sweater with water. (LSA value: 0.55)
- (Glenberg and Robertson calculated the coherence of meaning by using the LSA application (empowered by a large scale corpus) which is available at <u>http://lsa.colorado.edu/</u>)

But "Meshing" Affordance...

- I→Afforded ; II→Non Afforded
- "A sentence is meaningful to a particular reader to the extent that the reader can mesh the objects and activities as directed by the sentence" (Glenberg & Robertson, Symbol Grounding and Meaning, p.384)
- The meaning tells a best matching between body (physical) conditions and situation / environment conditions.
- The computer deprived of body turns out to be incapable of doing inference on the right usage of instruments.

Reflection

- Does the lack of affordance mean a simple nonsense (gag)?
- A "nonsense" sentence would presuppose a hidden context of meanings which is dangling out in spite of its incongruity and its extravagance.
- However, the argument of *Glenberg et al.* defines the affordance as the very first moment (or generation) of meaning without considering any semantic network which might be stealthily underlaid for all the possibilities of outcast meanings.



Is there no way to simulate such inhibitory relationship?

This is not the end of the war.

• Some day, the computer science will allow us to simulate this kind of subtle nonsense far beyond the limits of the vector space model.

 $- \rightarrow$ By the graph theory applied to a free association ?



Co-occurrence \rightarrow Association

- A type of association : Free Association
 - Free association is a process of generating ideas for writing through which one thought leads randomly to another.
 - <u>http://www.pearsoned.ca/text/flachmann4/gloss_iframe.html</u>
 - a thought process in which ideas (words or images) suggest other ideas in a sequence
 - <u>http://www.cogsci.princeton.edu/cgi-bin/webwn</u>
 - Free association (Psychodynamic theory) is a technique used in psychology, devised by Sigmund Freud.
 - <u>http://en.wikipedia.org/wiki/Free_association_(psychology)</u>
 - (Retrieved by using Google)
- How we can treat the free association:
 - Free, but grounded
 - Ex. Co-occurrence coefficient : a sort of correlation coefficient taking values from -1 to +1

Co-occurrence coefficient

(From Matsuo, *Meanings of the Word "Peace"* —*by Factor Analysis of Free Association Data*)

(Concrete research method) N participants are asked to present within a fixed time all the possible words thought of as related to a key word (for example, "peace").

	Word X	Word Y
Participant 1	1	0
Participant 2	0	0
Participant 3	1	1
Participant 4	0	1
Participant 5	1	0
Frequency	FX=3	FY=2
co-occurrence	CXY=1	
	1:presence	0:absence

Frequency of co-occurrence CXY means: how many participants present a particular set of 2 words at the same time

Co-occurrence coefficient *C* of the Words X and Y

$$C = 2\sum_{i=m}^{n} \left\{ \frac{F_{X}!(N - F_{X})!F_{Y}!(N - F_{Y})!}{i!(F_{X} - i)!(F_{Y} - i)!(N - F_{X} - F_{Y} + i)!N!} \right\} - 1$$

$$m = \max(0, F_{X} + F_{Y} - N)$$

$$n = \max(m, C_{XY} - 1)$$

N =Num of Participants F_X = Freq of X F_Y = Freq of Y C_{XY} =Freq of Co-occurrence $-1 \le C \le +1$

Factor Analysis of C data Matrix

- 5 factors were extracted by Prof. Matsuo from this free association data.
 - Universality-Individuality
 - Public concept Private mentality
 - Peacefulness, Quietness
 - Nature Artifact
 - Tendency of anti-militarization

Summary

- Eternal battle between the amodal theory and the embodiment theory
- There remains a slight hope for a computer simulation of affordance.
- Large-scale semantic network made by free association data would lead us to the resolution of the symbol-grounding problem.
- One way to treat a lexical co-occurrence data provided by a free association task: Co-occurrence coefficient.