Evaluating Semantic Relations and Distances in the Associative Concept Dictionary using NIRS-imaging

Nao Tatsumi[1], Jun Okamoto[2], Shun Ishizaki[1]
[1] Graduate School of Media and Governance, Keio University
[2] Keio Research Institute at SFC
[1,2] 5322 Endo, Fujisawa-city, Kanagawa, 252-8520, Japan
E-mail: nt@sfc.keio.ac.jp, juno@sfc.keio.ac.jp, ishizaki@sfc.keio.ac.jp

Abstract

In this study, we extracted brain activities related to semantic relations and distances to improve the precision of distance calculation among concepts in the Associated Concept Dictionary (ACD). For the experiments, we used a multi-channel Near-infrared Spectroscopy (NIRS) device to measure the response properties of the changes in hemoglobin concentration during word-concept association tasks. The experiments' stimuli were selected from pairs of stimulus words and associated words in the ACD and presented in the form of a visual stimulation to the subjects. In our experiments, we obtained subject response data and brain activation data in Broca's area －a human brain region that is active in linguistic/word-concept decision tasks－ and these data imply relations with the length of associative distance. This study showed that it was possible to connect brain activities to the semantic relation among concepts, and that it would improve the method for concept distance calculation in order to build a more human-like ontology model.

1. Introduction

Semantic relations between words are essential in language understanding for both computers and the human brain. The relation between words is the key to understanding the contents in the text as well as its syntactic or shallow semantic information. For natural language processing, this background knowledge of words is often represented in linguistic dictionaries or ontology. There are several large databases describing the relations between words. "The Nihongo GoiTaikei" developed by the Nippon Telegraph and Telephone Corporation (NTT) contains 400,000 Japanese words, and "The EDR Electric Dictionary" maintained by the National Institute of Information and Communications Technology (NiCT) contains 410,000 concepts.

The Associative Concept Dictionary (hereinafter ACD) is a dictionary with a hierarchical structure of words associated by human subjects with about 1,000 stimulus words on the basis of the following seven relationships: hypernym, hyponym, part/material, attribute, synonym, action, and situation. The ACD has a unique characteristic because it includes associative distance information or a kind of familiarity between two words, while the GoiTaikei and EDR do not include word frequency or other significant information. With this distance information, the ACD has been widely used with neural network models, for example Okamoto (2007) proposed a system to elucidate word sense disambiguation.

However, this distance is approximately calculated using a linear programming method, and the method does not use a parameter of the association time in the association experiments, which is related to human language cognition.

In the previous studies have shown that different spatial patterns of neural activation are associated with thinking about different semantic categories of pictures and words; for example, tools, buildings, and animals, and Mitchell (2008) showed the correlation between meaning of words from the word text corpus and its brain imaging data.

In this study, we will propose a method to improve distance calculation by using human cognitive and brain activity features. This method improves the conventional distance calculation method, and the improved distance will be closer to the human cognitive distances among concepts.

2. Associative Concept Dictionary

The Associative Concept Dictionary has been built on the basis of the results of large-scale online association experiments. In these experiments, the stimulus words are nouns obtained from Japanese elementary school textbooks. The subjects were requested to associate words from the stimulus words, with a given set of semantic relations: hypernym, hyponym, part/material, attribute, synonym, action, and situation.

The ACD has been constructed by using all of the associated concepts that are connected to the stimulus words, with the distances calculated using a linear programming method (Okamoto 2001). The ACD is built using quantified distances and is organized in a hierarchical structure in terms of the hypernym and hyponym. Attribute information is used to explain the features of the given word. In the association experiment, each stimulus each stimulus word had 50 subjects, who were students from the Shonan Fujisawa Campus of Keio University. The number of stimulus words is currently 1,100. The total number of associated words is about 280,000 and if excluding the overlapping words, it will be about 64,000.

The Figure 1 shows a concept dictionary description. In the figure, the word "chair" is a stimulus word for the association experiment and "furniture" is a higher-level concept of "chair." The numbers in the figure indicate individual associative distances.

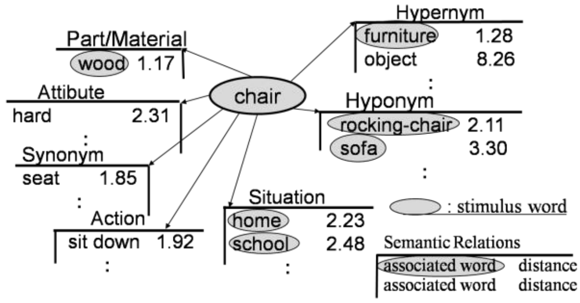

Figure 1: Concept dictionary description for the stimulus word "chair" (originally in Japanese)

3. Evaluate ACD using NIRS

To evaluate associative concepts and their distances, we use an NIRS-imaging device to measure oxygen assumption in the parts of the brain that are related to language cognition.

3.1 Method: NIRS

Near-infrared Spectroscopy (NIRS) is an optical imaging method that allows a non-invasive measurement of the changes in the concentration of oxygenated and deoxygenated hemoglobin in cerebral vessels of the brain or other tissues. NIRS uses near-infrared light between 650 and 950 nm and can provide quantitative measures of the hemoglobin changes that can help resolve ambiguities in the effects of blood flow and oxygen consumption responses to brain activation.

The NIRS-imaging device has the advantage of being a portable and inexpensive device for functional neuroimaging when compared to other brain imaging methods, like fMRI. Moreover, it does not require severely limiting subject motion. NIRS has well-balanced resolutions. It can measure these hemodynamic signals with a temporal resolution of 25 ms at the minimum, significantly greater than fMRI, aiding in the resolution of the onset time of brain activation. Spatial resolution is normally about 30 mm, but this can be improved by measurement probe alignment. In recent years, NIRS has been used to study various human cognitive functions, such as auditory and visual language processing.

3.2 Experiment

3.2.1. Subject

The subjects comprised nine healthy young adults (five males and four females) between the ages of 20 to 28. All subjects were right-handed and Japanese. Before measurement, the subjects were provided with a written informed consent form after receiving a full explanation of the study. All subjects have participated in association experiments for ACD and understand the associative concepts and their functions.

3.2.2. Record

For measurement, we used a multi-channel NIRS-imaging device (FOIRE OMM-3000/16, Shimadzu, Japan). The sampling time for the data acquisition was 40 ms. The measurement point intervals were 9 mm, by using a strip of plastic sheet of a fixed size, which was in accordance with the selectivity alignment method (Kato, 2004). The probes were placed on bilateral areas and covered from BA45 to the posterior of BA22. These are the well-known regions of the brain that are linked to various language functions.

3.2.3. Experiment System

We developed an electric device and software to synchronize an NIRS-imaging device with a computer, which the subjects used during the experiment (as shown in Figure: 2). The software can automatically present stimuli to the subjects on the computer display (19 inch) and record the subjects' mouse-click logs (in Experiment 2 only). The software can also send signals to the electric device to mark up the measurement data linked to the task. We marked (1) stimuli start, (2) stimuli end, and (3) subject's mouse click, to help in the data analysis.

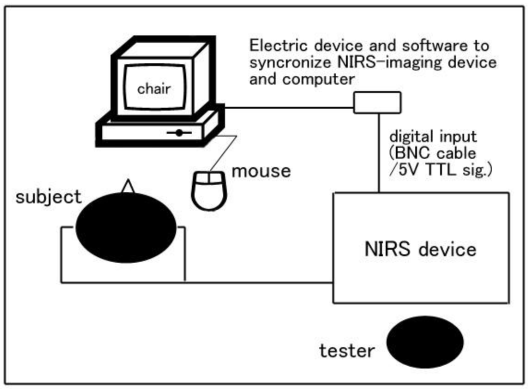

Figure 2: Experiment system

3.2.2. Analysis

We examined the subjects' responses (the response time in Experiment 2) and the response properties of the changes in the hemoglobin concentration. The time of stimuli presentation was referred to as 0 second (baseline), and the lapse of time from the baseline was calculated for each trial. In the hemoglobin analyses, we evaluated the cerebral deoxygenation or oxygenation in the capillaries in each measurement channel. We compared data sets by using a t-test—task (each relation) and distance (Long group/Short group)—Tukey's HSD, and an ANOVA.

3.2.2. Task

In this experiment, we conducted two tasks. In Experiment 1, we investigated the correlation between associative concepts/distances and human brain activation. Then, in Experiment 2, we conducted a concept judgment task to investigate how the response times and accuracy rates of the tasks were correlated to the associative distance data.

4. Experiment and result

4.1 Experiment 1: response to words relation

The subjects were instructed to look at visual stimuli and then to rest for less than three seconds between each trial. The stimuli were Japanese word-pairs, and we presented 300 trials in total. All the stimuli were selected from ACD and presented randomly. The word-pairs consist of stimulus words and their associate words from ACD, in different concepts and distances.

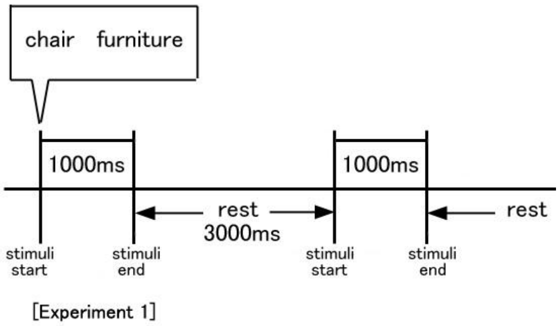

Figure 3: Task protocol for Experiment 1

4.2 Results of Experiment 1

Brain activation showed different patterns in response to different concepts. For example, during the action of concept stimuli, deoxygenation of hemoglobin occurred in the left Broadmann's Area (BA) 22 and in the right posterior of BA45 to 44. The right BA45 was also activated by part/material concepts; thus, it suggests that this response is related to word image processing. In addition, we would like to emphasize that we demonstrated the relationship between associative distance and brain activation in BA44. This experiment clarified that it is possible to correlate brain activation with associative relations among concepts.

4.3 Experiment 2: associative concept judgment

The subjects were instructed to look at visual stimuli and perform associative concept judgment tasks by mouse clicking, followed by a rest period of less than seven seconds between each trial. The stimuli were Japanese pair words, and we presented 70 trials. All the stimuli were selected from the ACD and presented randomly.

The task was associative concept judgment. The subjects were expected to look at three words—the first being a stimulus word; the second, a semantic (associative) relation; and the last, an associative word in the ACD—and judge whether stimulus-associate word correspondence is true or false by mouse clicking. For example, if a trial presents the stimuli [chair–hypernym–furniture], the answer will be "true" if it presents [dog–part/material–wing], the answer will be "false."

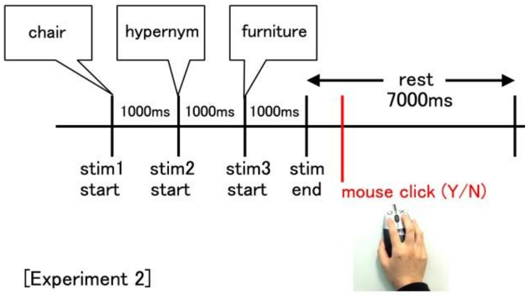

Figure 4: Task protocol for Experiment 2

4.4 Results for Experiment 2

We are working on the NIRS data analysis for Experiment 2. Until now, we have obtain the subjects' response times and correct rates for the associative concept judgment task. The data on timings and rates showed differences with the various kinds of stimuli, that is, a kind of semantic relation or the length of associative distance.

In a few samples, these differences imply a relation to the length of associative distance and human brain activation.

5. Conclusion

In our study, we obtained subject response data and brain activation data related to associative relations and distance. NIRS can provide quantitative measures of human cognition, which are indicated in the changes in hemoglobin concentration in response to brain activation. These can be considered as parameters for recalculating associative distances.

We are continuing an NIRS data analysis to conduct further results and may employ them to validate and improve associative distance calculation in the ACD.

6. Acknowledgements

This study has been supported by the Mori Memorial Research Fund (Graduate Student Researcher Development 2009-No.7) at Keio University. We would like to thank Ms. Kayoko Yoshino for her NIRS technical advice. We would also like to extend our gratitude to the students who were the subjects of our NIRS experiments, and to the students who participated in the association experiments for the ACD at SFC, Keio University.

7. References

Gabrieli et al. (1998), The role of left prefrontal cortex in language and memory. PNAS 95: 906-913.

Kato, T. (2004), Principle and technique of NIRS-Imaging for human brain FORCE. International Congress Series 1270C: 88-99.

Mitchell, T et al. (2008), Predicting Human Brain Activity Associated with the Meanings of Nouns, Science 320, pp.1191-1195.

Okamoto, J. & Ishizaki, S. (2001), Construction of Associative Concept Dictionary with Distance Information, and Comparison with Electronic Concept Dictionary (in Japanese). Journal of Natural Language Processing 8(4): 37-54.

Okamoto, J. and Ishizaki, S. (2007), Word Sense Disambiguation on Contextual Dynamic Network Using Associative Concept Dictionary, PACLING 2007, pp. 93-100.