Linguistics D: Class 7 Is a language for messages, or for something else?

1 Research project/Conference preparation

Q.1 What is the point you should emphasize most in your research presentation? Where should it be written in your poster.

Q.2 After the conference, organize into in a commentary page, A4, 1 page, words you have used in your presentation, in your colleagues's comments and questions, and submit your poster and the commentary together (two pages in total) from the link of this page.

2 Theoretical aspect: Is a language for messages, or for something else?

In general, we will use language to tell something, probably a message, to somebody. However, sometimes, we will use language for nothing. Or, there are some scenes where it is difficult to find out the reason why we use language.

Q.3 What is the purpose of chatting at mealtime? For the purpose of what, will you have a chat at mealtime?

Q.4 If the purpose exits, describe it. Or it does not, discuss why you will have a chat or make a conversation even if you do not have any purposes.

Q.5 Discuss whether the nature of language is to make a human relationship or to convey a message.

Q.6 Examine if there are some relating articles in textbooks or research archives with term "phatic communion." (Malinowski 1989)

3 Mathematical aspect: Weighting of a word

Q.7 Give examples of common and rarely seen words.

Q.8 There are two types of word: a word appears everywhere, and another appears rarely. Discuss the characteristics of these words.

Q.9 Discuss how to distinguish common words from rare words.

Q.10 Discuss whether similar distinctions exist in other or your disciplines.

Q.11 Discuss what the following equation means.

$$tf(i, N) = \frac{n_i}{N}$$

where i is a token; n is the frequency of appearance of token i, N is the total number of words in an expression region such as a text, texts, lines, a page, pages, ...

Q.12 Discuss what the following equation means.(Manning and Schütze 1999, Robertson 2004, Spärck Jones 1972)

$$tfidf(i, D) = tf(i, N) \cdot idf(i)$$

$$idf(i) = log\frac{D}{d_i}$$

where d is the number of document where token i appears; and D is the total number of expression region.

Q.13 Discuss how $tfidf(i, D)$ will distribute.

Q.14 We have two tokens which co-occur in some expression regions. Discuss if it is possible to calculate the following expression instead of $tfidf(i, D)$. And discuss the distribution type/shape of cw as well.(Gross et al. 2012, Yamamoto and Hodošček 2018)

$$cw(i_1, i_2, D) = ctf(i_1, i_2, N) \cdot cidf(i_1, i_2)$$

4   Topic of the day: Linguistic Distance

Q.15 Discuss what language is closest to English. Also, discuss what language is closest to Japanese.

Q.16 Discuss how you define the proximity between two language?

Q.17 Discuss how to quantify the distance between two words.

5   Sentence structure of Japanese: Meaningless/nonsense words

Q.18 Give examples of words that are hard to tell their meaning.

Q.19 Discuss whether words that are difficult to tell their meaning are really meaningless.

Q.20 Discuss why such words exist.

6   Further exercises

Q.21 Find some articles relating to today's questions/topics in textbooks your instructor introduced, and describe the name of the book, the number of pages, the title of the article, and your opinion.

7   Homework

Q.22 Access the web page from the QR code and answer the questions (deadline: today).

References

Gross, Oskar, Hannu Toivonen, Jukka Toivanen, and Alessandro Valitutti (2012) "Lexical Creativity from Word Associations".

Malinowski, Bronislaw (1989) "The Problem of Meaning in Primitive Languages", in C. K. Ogden and I. A. Richards eds. *The Meaning of Meaning*, San Diego, New York, and London: Harcourt Brace Jovanovich, Chap. Supplement 1, pp. 296–336.

Manning, Christopher D. and Hinrich Schütze (1999) *Foundation of statistical natural language processing*, Cambridge, Massachusetts: The MIT press.

Robertson, Stephen (2004) "Understanding inverse document frequency: on theoretical arguments for IDF", *Journal of Documentation*, Vol. 60, pp. 503–520.

Spärck Jones, Karen (1972) "A Statistical Interpretation of Term Specificity and Its Application in Retrieval", *Journal of Documentation*, Vol. 28, pp. 11–21.

Yamamoto, Hilofumi and Bor Hodošček (2018) "A study on the distribution of cooccurrence weight patterns of classical Japanese poetic vocabulary", *JADH2018 Proceedings of the 8th Conference of Japanese Association for Digital Humanities "Leveraging Open Data"*, Vol. 2018.

Homework submission