語意，岐義，與多義：以實證方法界定漢語語意的分界 （II）
Sense, Ambiguity and Polysemy: Empirical Approaches to Delimiting Senses – Part II

中文摘要

此份報告內有「語意，岐義與多義：以實證方法界定漢語語意的分界」兩年來的研究成果。本研究主要目標為以基於詞料庫和心理語言學實驗的兩種研究方法來決定岐義名詞（多義的名詞）的詞彙表徵。

關鍵字：詞彙表徵，岐義，語料庫語言學，心理語言學

Abstract

This paper reports on the results of the research conducted for the two-year grant 'Sense, Ambiguity, and Polysemy: Empirical Approaches to Delimiting Senses.' The main goal of this research is to determine the lexical representation of ambiguous nouns (nouns that have multiple meanings) with the aid of both corpus-based approaches as well as experimentally-based psycholinguistic approaches.

Keywords: lexical representation, ambiguity, corpus-linguistics, psycholinguistics

Background and Goals

In this two year research study, I asked the following questions:

1. How should nominal meaning be best represented (i.e. how should words with multiple meanings be represented?)

2. Are words with many senses accessed faster than words with fewer senses (i.e. does the number of meaning (NOM) effect exist)?

3. Does sentential context influence lexical access when a word has more than one meaning?

The first question relates to fundamental issues in representation, such as how to deal with meaning extension instances of metaphor and metonymy, and when such meanings extensions should be considered to involve a new 'sense'. The second question hinges on defining exactly what a sense is, and as such is closely related to the conclusions we reach concerning question 1 above. It has implications for understanding the nature of on-line lexical access. The third question relates to the issue of modularity. If context does influence lexical access of ambiguous words, then that would mean that the lexical module is not insularized from the pragmatic-semantic module, which would argue for an interactionist view of sentence processing. In what follows, I will briefly answer each question in turn, referring the reader to papers published as a result of this research for more in-depth analyses and discussion.

Nominal Representation

Question 1: How should nominal meaning be best represented (i.e. how should words with multiple meanings be represented?)

I propose that there are two levels of meaning representation: **senses** and

meaning facets (Ahrens et al. 1998). A slightly revised and updated definition of sense and meaning facet is given below:

Meaning facets, on the other hand, have the following distinguishing properties:
1) are an instance of metonymic or meronymic extension;
2) nouns of the same semantic class will have similar extension links to related meaning facets.
3) can appear in the same context as other meaning facets (Note: this condition is weakened in the cases of mono-syllabic words);

Senses have the following properties:
1) are not an instance of metonymic or meronymic extension, but may be an instance of metaphorical extension;
2) the extension links between two senses cannot be inherited by a class of nouns.
3) cannot appear in the same context (unless the complexity is triggered);

There are two basic types of extensions for meaning facets: meronymic and metonymic extensions. Meronymic extensions involve both the whole standing for the part and the part standing for the whole, and are driven by cognitive and conceptual saliency. Metonymic extensions are typically driven by eventive relationships, such as the information media standing for the information creator, or a container standing for a containee, or an event standing for a temporal period. Please refer to Ahrens et al. (1998) for a further discussion of these extensions.

Multiple senses may be a result of metaphoric extension, although they do not have to be. A metaphorical usage of a word is an example of a different sense because a metaphorical meaning must appear in a different context, and because the link between the two senses cannot be inherited by a class of nouns. We make a distinction between novel metaphorical senses and conventionalized metaphorical senses, including only the latter in our representation.

The main advantage of our representation is that it allows for the fact that meaning may be actively complex, which means that simultaneous multiple interpretations of a word exist. That is, whereas previous models of lexical ambiguity resolution assume that only one solution exists in a given context, our account allows for more than one meaning to co-exist in a particular context. Other advantages for this representation are also given in Ahrens et al. (1998).

To date, we have analyzed the meanings of 225 nouns in 27 semantic categories. All nouns analyzed to date are either inanimate or animate concrete nouns. Our method is as follows: we search the corpus for each word, and then print out all the sentences that the word occurs in. Then we analyze the meaning of the word in each sentence. After this semantic analysis, which is checked by another native speaker. the word's meaning (including its senses and meaning facets) is listed along with all relevant examples. These entries are entered into a database which will allow searches for words with multiple senses and meaning facets. The database also allows for searches of words of a particular semantic class. Furthermore, when one clicks on a sense or meaning facet, it is also possible to access all examples of that particular sense or meaning facet. We hope this database will be useful to linguists, computational linguists, and psycholinguists in the future.

In addition to the question of how nominal meaning should be represented, I have also started looking at how meaning is mapped in metaphorical extensions between the source and target domains (Ahrens & Say 1999). Metaphorical extension of meaning in Chinese and its associated principles is one area for future research in terms of both nominal representation as well as processing of meanings.

Number of Meaning Effect

Question 2: Are words with many senses accessed faster than words with fewer senses (i.e. does the number of meaning (NOM) effect exist)?

Rubenstein et al. (1971) found that words with more meanings were accessed faster than words with fewer meanings. However, subsequent studies have not always supported this finding. However, upon examining the literature, we find that previous researchers did not have a linguistic definition of 'sense.' They instead relied on either native speaker judgments or information in a dictionary. I conclude that the reason that findings exist both for and against the NOM effect is precisely because the notion of what constitutes a 'sense' is not clearly delimited. In our study we define sense according to the linguistic criteria given in Ahrens et al. (1998). We hypothesize that when this linguistic notion of sense is used, the NOM effect will be shown to be robust. The implication of our hypothesis is that words with more senses are accessed faster because they take up more space in the mental lexicon, that is, it is easier for the processor to locate and entry that is taking up more room. This is known as the Random Access Model (RAM). (The experimental work on this section of the grant was carried out by Charles Lin, as part of his Master's thesis at the Graduate Institute of Linguistics, National Cheng Chi University, under the direction of myself and Professor I-Li Yang).

For this experiment, we collected the meanings of 200 nominals from 200 Mandarin-speaking college students and analyzed their number of senses according to Ahrens et al. (1998), discussed in Section 1 above. We looked at three possible factors which could affect rate of lexical access. First we looked at the number of senses a word had. Second, we looked at meaning relatedness, which predicts that words will more closely related meanings will be accessed faster.

Third, we looked at the distribution of meaning frequency, with the idea that the more equally frequent the senses of a word are, the faster a word will be accessed. When we looked at each of these factors, we controlled for the other two factors, in addition to controlling for character and word frequency, as well as character complexity and word familiarity. The main task was for subjects to look at a computer screen and to decide if the two-character word that flashed on the screen was a word or not. They pressed the non-word button if it was not a word, and the word button if it was a word. Then their correct responses were averaged over condition and submitted to statistical analyses. We found that words with more senses were accessed faster than words with fewer senses, which is what Rubenstein et al. (1971) originally predicted. However, we did not find evidence for meaning relatedness or distribution of meaning frequency. This finding is important because it demonstrates that a word's number of senses is the most likely metric that the language processor holds on to when looking to access a word. This concept is also, when compared with the other two, the most salient and easily calculable number. This finding also argues against a serial account model of lexical access, and instead supports the random access model. That is, the more senses a word has, the faster that one of those senses will be accessed. Lastly, because the number of senses calculated was based on the representation suggested in Ahrens et al. (1998), it lends plausibility to that representational model from a processing point of view.

Effect of Context on Lexical Access

Question 3. Does context influence lexical access when a word has more than one meaning?

The purpose of the experiment presented in this experiment is to determine if the lexical access of ambiguous words (words with two senses)

are influenced by preceding sentential context. One hypothesis, known as the interactionist hypothesis, postulates that preceding sentential context can influence lexical access of an ambiguous word in on-going sentence comprehension. Another hypothesis, the modularity hypothesis, postulates that preceding sentential context cannot influence lexical access of an ambiguous word in on-going sentence comprehension. Within these two frameworks there are a range of models. Please see Ahrens (1998) for a discussion of these two hypotheses and sub-models.

Two experiments test specifically if lexical access is influenced when the sentential context is biased towards the secondary meaning of the ambiguity. We present the visual probe at the immediate offset of the lexical ambiguity and present it for 300ms in the first experiment and for 1500ms in the second experiment. Our hypothesis is that the short presentation time will demonstrate that both meanings are first accessed, but the longer presentation time will give context the chance to make a choice, and only the contextually appropriate meaning will be accessed. This finding would be in line with the modularity hypothesis, which predicts that context cannot influence *immediate and automatic* lexical access, but that once immediate and automatic access has taken place, context can come in and make a selection. This is, in fact, what we find.

Our results demonstrate that when the length of the presentation probe is 1500ms, only the contextually appropriate meaning is still available (Ahrens 1999a,b). However, since both meanings were available at the offset of the ambiguity when the visual probe was presented for 300ms, the fact that only the contextually appropriate meaning was available at 1500ms indicates that the non-contextually appropriate was initially available, but then faded. This is precisely what the modularity hypothesis predicts.

Conclusion

In sum, over the past two years we have suggested a new model for representation of multiple senses, analyzed over 200 nouns based on corpus data within this model, created a database for other linguists to use that accesses our analyses and the original data. We have also run three on-line psycholinguistic experiments. The first one was a straightforward lexical decision task that demonstrated that Rubenstein's original (1971) proposal that words with more senses are accessed more quickly than words with fewer senses, when all other factors are held constant. This finding supports a random access model of lexical access, as opposed to a serial model of lexical access. In addition, since the definition of sense was based on our linguistic definition (Ahrens et al. 1998) it demonstrates that this definition is useful in evaluating the number of senses that a word has for a particular group of speakers. Lastly, we also ran two cross-modal lexical decision task experiments which demonstrated that context does not influence immediate and automatic lexical access. However, once immediate and automatic lexical access has taken place, context does help the processor select the contextually appropriate meaning. This finding supports the modularity hypothesis in language processing.

References

Ahrens, Kathleen. 1999a. Timing Issues in Lexical Ambiguity Resolution. Presented at the North American Conference on Chinese Linguistics, Harvard University. June 18-20.

Ahrens, Kathleen. 1999b. Timing Issues in On-line Sentence Processing. International East Asian Psycholinguistics Workshop. Ohio State University. August 3-4.

Ahrens, Kathleen. 1999. On-Line Sentence Comprehension of Ambiguous Verbs in Mandarin Chinese. *Journal of East Asia Linguistics*.

Ahrens, Kathleen. 1998. Lexical Ambiguity Resolution: Languages, Tasks and Timing. In *Sentence Processing: A Cross-linguistic Perspective*. (Ed.) Dieter Hillert. Academic Press. Pp. 11-31.

Ahrens, Kathleen, Li-li Chang, Keh-jiann Chen, and Chu-Ren Huang. 1998. Meaning Representation and Meaning Instantiation for Chinese Nominals. *Computational Linguistics and Chinese Language Processing* 3, 45-60.

Ahrens, Kathleen and Alicia Say. 1999. Mapping Image-Schemas and Translating Metaphors. *Proceedings of the 13th Pacific Asia Conference on Language, Information, and Computation.*

Rubenstein, H., S. Lewis, and M.A. Rubenstein. 1971. Homographic Entries in the Internal Lexicon: Effects of Systematicity and Relative Frequency of Meanings. *Journal of Verbal Learning and Verbal Behavior* 10, 57-62.