# A Language Analysis Method for Spoken Dialogue Understanding

**Manabu OKUMURA**
School of Information Science,
Japan Advanced Institute of Science and Technology
Tatsunokuchi, Ishikawa 923-12, Japan
e-mail: oku@jaist.ac.jp

## 1. Introduction

Short words often fail to be recognized in speech recognition. This situation holds true for Japanese postpositions. In traditional semantic analysis for Japanese, postpositions are often used as a key for selecting the appropriate case frame. However, in semantic analysis for spoken utterances, since postpositions are often unrecognizable, it is difficult to apply the traditional approach.

In this paper, we present a semantic analysis method for spoken dialogue understanding that uses only the content words in utterances, that is, without the use of postpositions. Our work can be divided into two parts:

1. Language analysis based on lexical cohesion,
2. Case frame selection with the cooccurrence information of nouns and verbs from corpora.

First, we describe how word sense ambiguity can be incrementally resolved with the aid of lexical cohesion. Then, we describe how segment boundaries of a discourse can be determined with the aid of lexical cohesion.

Secondly, we describe a case frame selection method only with the information of nouns and verbs that uses the cooccurrence information from corpora.

## 2. Lexical Cohesion

A discourse is not a mere set of unrelated sentences. Rather, sentences in a discourse are about the same thing and connected to each other. Lexical cohesion is said to contribute to such connection of the sentences. We call a sequence of words which are in lexical cohesion relation with each other a lexical chain. Lexical chains tend to indicate portions of a discourse that form a semantic unit. Therefore, lexical chains provide a local context to aid in the resolution of word sense ambiguity. Here we use a Japanese thesaurus 'Bunrui-goihyo and we count a sequence of words which are included in the same category as a lexical chain.

### 2.1. Incremental Word Sense Disambiguation Based on Lexical Cohesion

During the process of generating lexical chains incrementally, they are recorded in a register in the order of the salience. The salience of lexical chains is based on their recency and length. Since the more salient lexical chain represents the nearby local context, by checking lexical cohesion between the current word and lexical chains in the order of the salience, in tandem with generation of lexical chains, we realize incremental word sense disambiguation based on contextual information that lexical chains reveal.

We apply the algorithm to real discourses. The average performance is 63.4%. We think the system's performance is promising for the following reasons: Lexical cohesion is not the only knowledge source for word sense disambiguation and proves to be useful at least as a source supplementary to the framework that uses case frames.

### 2.2. Discourse Segmentation Based on Lexical Cohesion

As mentioned above, lexical chains tend to indicate portions of a discourse that form a semantic unit. Therefore, lexical chains provide also a clue for the determination of segment boundaries of the discourse. We think it is crucial to identify the segment boundaries as a first step to construct the structure of a discourse.

When a portion of a discourse forms a semantic unit, there is a tendency for related words to be used. Therefore, if lexical chains can be found, they will tend to indicate the segment boundaries of the discourse. When a lexical chain ends, there is a tendency for a segment to end. If a new chain begins, this might be an indication that a new segment has begun. Taking into account this correspondence of lexical chain boundaries to segment boundaries, we measure the plausibility of each point in the discourse as a segment boundary: for each point between sentences $n$ and $n+1$(where n ranges from 1 to the number of sentences in the discourse minus 1), compute the sum of the number of lexical chains that end at the sentence $n$ and the number of lexical chains that begin at the sentence $n+1$. We call this naive measure of a degree of agreement of the start and end points of lexical chains $w(n, n+1)$ *boundary strength*. The points in the discourse are selected in the order of boundary strength as candidates of segment boundaries. Furthermore, since lexical chains have gaps inside themselves, we refine the measure by taking into account their gaps. The start and end points of the gaps also can get the boundary strength.

To evaluate our system, we pick out 22 discourses, which are from the questions of the Japanese language that ask us to partition the discourses into a given number of segments. The system's performance is judged by the comparison with segment boundaries marked as an attached model answer. Since human subjects do not always agree with each other on segmentation, our evaluation method using the discourses in the questions with model answers is considered to be a good simplification. The system's performance is evaluated in two ways. One is in case where the system generates the given number of segment boundaries in the order of the strength. From this evaluation, we can compute the system's marks as an examinee in the test that consists of the questions. The other is in case where segment boundaries are generated down to half of the maximum strength. In this case we use the following metrics for the evaluation: *Recall* is the quotient of the number of correctly identified boundaries by the total number of correct boundaries. *Precision* is the quotient of the number of correctly identified boundaries by the number of generated boundaries.

Here we do not take into account the information of paragraph boundaries, such as the indentation, at all in the following reasons:

- Because our discourses are from the exam questions, many of them have no marks of paragraph boundaries;
- In case of Japanese, it is pointed out that paragraph and segment boundaries do not always coincide with each other.

## 3. Case frame selection with the cooccurrence information

The triplets $< noun, deepcase, verb >$ can be extracted from word-sense tagged corpora as the cooccurrence information. We compute the score for the combination of nouns and a verb in sentences that is based on the frequency of the triplets, and select the case frame that can get the highest score. To cope with the problem of data sparseness, the score is defined on classes in the thesaurus, not on words, as Resnik first proposed in 1992.

In fact, this 'class-based' approach is effective in many cases, in the case where the score can be computed only in quite upper classes, that is, in the case only with quite few data, the accuracy of the case frame selection might degrade, because the score does not reflect the information of the relationship between nouns and cases. To cope with the problem, we multiply the score by the weight that is calculated with the frequency of the tuples $< noun, deepcase >$.

We evaluate our method with 150 sentences from a speech database. We use the EDR cooccurrence dictionary and EDR corpora as the corpora in this work. The number of the deep cases that EDR dictionary adopts is 24. From the result of the experiments where five human subjects try to select the correct case for each test sentences from 24 candidates, we make the data for evaluating the performance of our method.

The result of the experiment with our method shows that weighting the score is effective and improves the accuracy of the case frame selection. However, it also becomes clear that adjusting the proportion of the weight to the score will improve the accuracy further.

Keywords: content words, lexical chain, incremental NLP, local context, contextual analysis, discourse structure, case frame selection, corpus-based, data sparseness problem