

An Example-Based English Learning Environment for Writing

Yasuo Miyoshi^a, Youji Ochi^a, Ryo Okamoto^b, Yoneo Yano^a

^a Faculty of Engineering, Tokushima University
2-1 Minamijosanjima, Tokushima, 770-8506, Japan
Tel: +81-88-656-7495, E-mail: {miyoshi, ochi, yano}@is.tokushima-u.ac.jp

^b Faculty of Science, Kochi University
2-5-1 Akebono, Kochi, 780-8520, Japan
Tel: +81-88-844-8341, E-mail: ryooka@is.kochi-u.ac.jp

Abstract

In writing learning as a second/foreign language, a learner has to acquire not only lexical and syntactical knowledge but also the skills to choose suitable words for content which s/he is interested in. A learning system should extrapolate learner's intention and give example phrases that concern with the content in order to support this on the system.

However, a learner cannot always represent a content of his/her desired phrase as inputs to the system. Therefore, the system should be equipped with a diagnosis function for learner's intention. Additionally, a system also should be equipped with an analysis function to score similarity between learner's intention and phrases which is stored in the system on both syntactic and idiomatic level in order to present appropriate example phrases to a learner.

In this paper, we propose architecture of an interactive support method for English writing learning which is based on analogical search technique of sample phrases from corpora. Our system can show a candidate of variation/next phrases to write and an analogous sentence that a learner wants to represent from corpora.

Keywords:

writing support; example retrieval; corpus; phrase structure

Introduction

In recent years, the opportunity where we touched English increased by the spread of the Internet. Furthermore, when we transmitted information towards the whole world, the ability that we represent our intention in English is needed. Accordingly, English writing learning comes to increase in importance more than before.

The skills for writing are acquired experientially, but it is difficult to obtain the sufficient learning effect by short training. Beginners almost train given expressions repeatedly, as a drill, because they cannot express their

thoughts in English. Therefore they may not be able to learn the expressions that they really need, and their interest for training may fade. An environment to train free writing is necessary for them.

Our proposing environment has the following feature. When a learner comes to a deadlock in the process of free writing, his/her liability is mitigated and the efficiency of writing is improved, by showing the examples, which are similar contents of his/her intention. Consequently, seeing the useful examples can expect the learning effect.

This paper is to propose a learning environment and a learning support method for learning experientially. To begin with, we focus attention on the method that learners write the foreign language by referring to existing documents, and we point out inadequacies of the existing writing support method. Next, we describe the writing support method that we have proposed for learning. Furthermore, we describe the prototype of the English writing support system now being developed.

Writing with Reference to Examples

When non-native speakers write in English, most of them use English dictionaries. But to hesitate about which word to select is the matter of common occurrence when more than one word existed in the dictionaries. Although a lot of examples are inserted in a popular dictionary, there is little probability of applying to the case that the examples match the writers' intention.

Then they refer from their favorite existing documents to English expressions of the similar contents of their intention, as Figure 1 indicated. At that time, they need to know where of the existing documents the useful expressions are written. Because to search the expressions one at a time by humans is too non-effective.

However, by using a computer, it is possible to search the useful expressions from documents that they have not read. Further they can know how many expressions, that they are

just about writing, exist practically in documents. This thing means that they can solve their doubts similar to occurring when they examined words by using dictionaries.

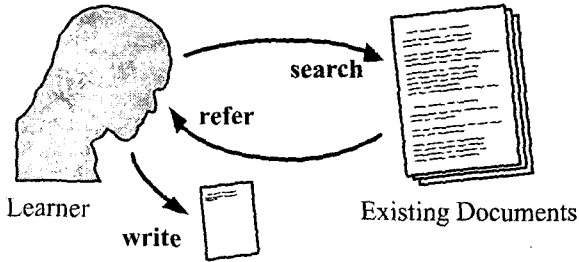


Figure 1 - writing by referring to existing documents

The many systems for searching similar examples already exist, and some of them are developed as a tool for acquiring composition ability[1][2]. When a learner writes with reference to examples using the retrieval system, as Figure 2 indicated, instead of a learner, the system searches the examples out of the corpora, which accumulate large quantity documents. Thereby, a learner can find the examples for referring to more efficiently. The system is useful to supporting training of writing learning by learner's independent demand.

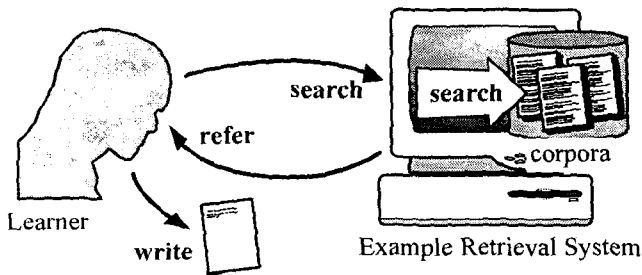


Figure 2 - Writing with "example retrieval system"

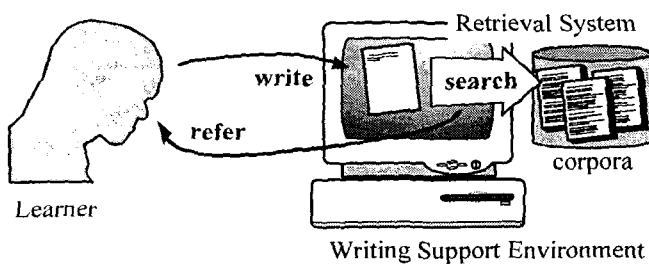


Figure 3 - Writing with "writing support environment"

However, their many systems perform very simple retrieval by character sequences which the learner inputted, a learner finds it hard to get satisfactory examples. Moreover, for a learner, to give complicated search conditions in order to obtain satisfactory examples is troublesome operation and excessive liability.

Then, we propose the English writing support environment with a function of examples reference, for abolishing the necessity of considering complicated search conditions. Although the following chapter describe this support

environment in detail, it can make English writing to a learner freely. If a learner requires, learner's intention will be extrapolated from the contents of the sentence in the middle of writing, and similar example retrieval will be performed according to his/her intention. In writing, using this environment as Figure 3 indicated, a learner is released from search operation and is able to concentrate. Therefore a learner can train efficiently the writing skill which is difficult to acquire in a short period.

Writing Support Environment

In this chapter, we suggest the English writing support environment intended for beginners.

Overview

The purpose of this support environment is to lead learners' incomplete English sentence to right complete sentence by support based on examples, and to get them to learn efficiently by experiencing a process of writing. The environment is provided to a learner in order to write freely, and it retrieves and presents the examples that took into consideration the learner's intention, which is extrapolated by national language analysis and dialogue with the learner.

It may become any easy to approach to write an English composition under the favor of the system that is able to retrieve example sentences by inputted keywords. The learners, however, don't search the example sentences of the very things. They search the phrases that they can just apply. They either do not think that just the same sentence exists in another documents. Therefore we adopt the method of representing the useful phrases for learners in this environment in writing.

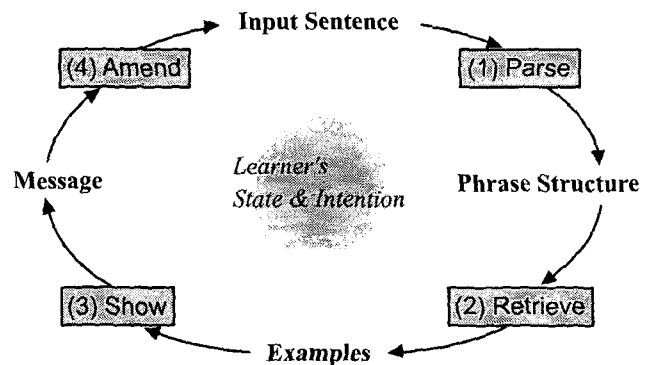


Figure 4 - Support Cycle for Writing

The cycle that supports English writing with our method is shown in Figure 4. As shown in the figure, the writing support is executed continually, as long as the demand from the learner continues.

At the first setout, a learner inputs some sentence. At that time, s/he is allowed to write only within his/her abilities. Then input sentence is parsed and added the tags expressing structure and parts of speech(1). Examples are searched with the tagging input sentence as search condition(2). The

support system of this environment represents the messages for reference from statistics of retrieved examples(3). At this point, when learner demanded support, the support system on the environment had to assess state of a learner from the input data. And the system have an insight into the learner's intention from his/her state.

Learner's state: The learner is in trouble because not to know suitable expressions, or is worrying whether there is any error in his/her expressions.

Learner's intention: The learner wants to know what expressions generally will be used, and/or wants to know how many his/her expressions will be used actually.

If learner's state is the former, his intention is to want to know what expressions will be used. Consequently, the system detects points to be not enough in his sentence, and represents phrases that can make up for the points. Such phrases are searched by analyzing the phrases similar to the phrases to be not enough. If learner's state is the latter, the system, as a simple example retrieval system, had better show how many sentences are similar to his sentence. Because it is possible that s/he feels some relief. The learner amends freely the sentence by referring to the represented message(4). Then, the learner can amend the sentence times out of number.

Architecture

This environment has an interactive interface for returning feedback to the natural language analysis results and the similar example retrieval results. The learner's feedback responses exert an influence on reflecting intention extrapolation. Because it is difficult to detect lame points in learner's sentence with the present natural language analysis technology. Figure 5 indicates the process flow that considered a response from a learner. The system decide the message presented to the learner by his /her response.

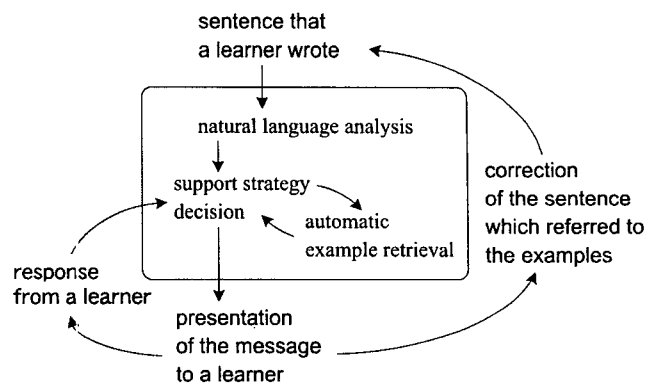


Figure 5 - Process Flow

Besides, the system considers to allow the input of Japanese for beginners that are not able to input anything. And implementation of dictionaries and thesauruses is necessary in order to show synonyms and paraphrases to a learner. The configuration of this environment to meet these demands is depicted in Figure 6.

Knowledge Base for Teaching

In the knowledge base, the useful knowledge for writing is stored. Dictionaries are natural language analysis dictionary and Japanese-English dictionary for referring to English words when input sentence contains Japanese words. Thesaurus has many synonyms. The documents stored in the corpus for retrieving the similar examples are altogether analyzed as hierarchical structure, and the corpus consists of tagged English sentences.

Learner Model

A learner model is essential to take an appropriate support action. A learner corpus scores the sentences written by the learner. Every action of a learner is stored in log data of dialogue.

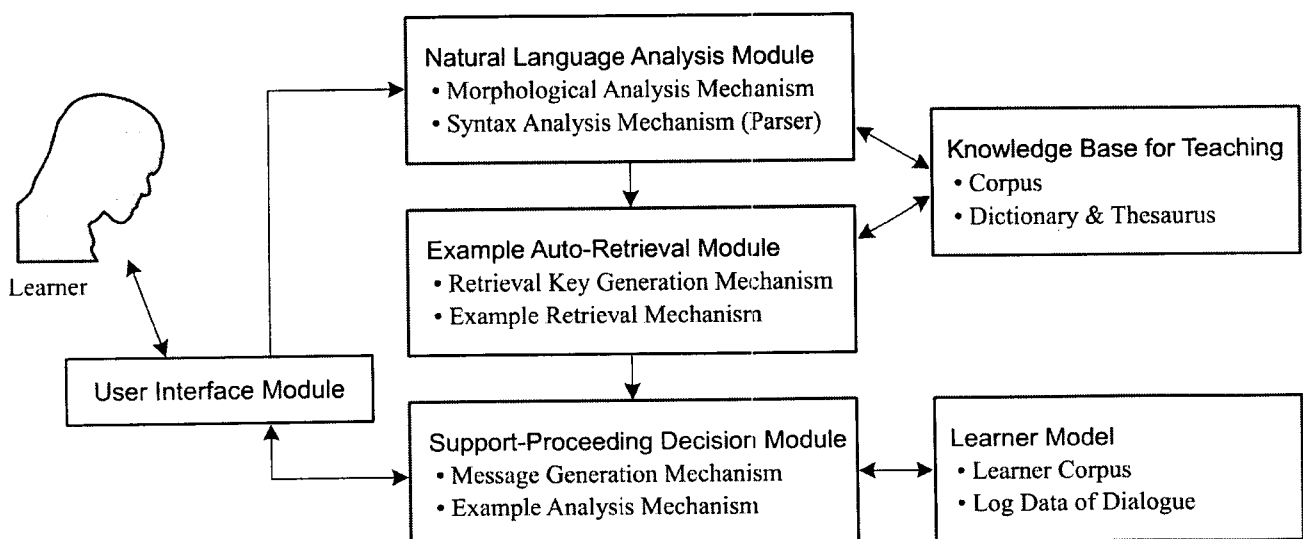


Figure 6 - Configuration of Support Environment

User Interface Module

User interface takes in the sentence that a learner made freely as the input and shows a message to a learner. As for the presentation of a message, when possible, the graphical representation is desirable for a learner. And this module accepts the demand a learner needs.

Natural Language Analysis Module

This module analyzes an input sentence of a learner and adds tags. When an input sentence includes Japanese, this module picks out only Japanese and divides it into a word. These Japanese words are translated to English words by using bilingual dictionary. We cannot expect a correct analysis result, because a lot of errors are included in the sentence that a learner input.

Example Auto-Retrieval Module

This module automatically retrieves the reference examples for writing. And this module consists of retrieval key generation mechanism and example retrieval mechanism. Retrieval key generation mechanism generates the key based on input sentences tagged. Example retrieval mechanism repeats example retrieval by changing a combination of the retrieval key. Both of data on keywords and tags are important to calculate the similarity in example retrieval.

Support-Proceeding Decision Module

This module consists of example analysis mechanism and message generation mechanism. Example analysis mechanism analyzes all retrieved examples, and gives those distinction. If there are not instructions from a learner, message generation mechanism assigns the support target points based on dialogue log data and fragmentary structure of incomplete input sentence.

Prototype System

We have developed the prototype of support system for writing that narrows down a target to the abstract of technical papers. The prototype system retrieves the similar sentences taking notice of the phrase structures of sentences, which are data inputted or examples of corpora.

Functions

The prototype system can represent following useful information to a learner.

- Related words: Synonyms of words in input sentences.
- Similar sentences: It has the phrase structures which match some phrase structures which constitute an input sentence. And it is sorted in order of similar degree.
- Variation/Following candidate phrases: It is candidate of variation/next about the phrases which matched the phrase of the last in an input sentence. And it is sorted in order of the frequency used by the corpora.

Configuration

Figure 7 shows the system configuration of the prototype system. The system runs under Linux except user interface program. The development language is C. User interface program runs under Mac OS. The development language of user interface program is REALbasic. The database server system is Postgres.

Libraries

This system uses four libraries. Three libraries except a libpq library do natural language analysis, and input sentence analysis program takes advantage of it.

- chasen: Library to use a module of Japanese morpheme analysis system 'Chasen'[3].
- wn: Library to access 'WordNet' used as thesaurus in the system[4].
- app: Library to use English sentence structure analysis system 'Apple Pie Parser'[5].
- libpq: Interface library to access PostgreSQL

Dictionaries

This system implements two dictionaries to allow input sentences including Japanese words. Data of a dictionary has moved to Postgres database.

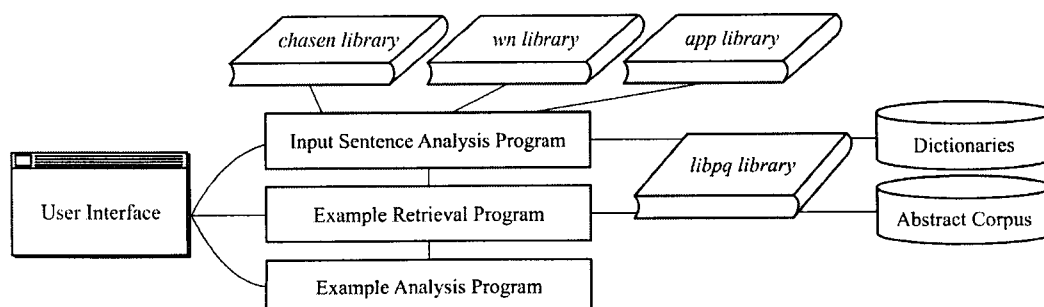


Figure 7 - Configuration of the Prototype System

- Japanese-English Dictionary: Kenkyusha 'Waeichujiten' CD-ROM. Registration total vocabulary are about 36,000.
- IPAL Dictionary: IPAL is Japanese morpheme dictionary for computer. This system uses IPAL as Japanese thesaurus. [6]

Abstract Corpus

The data of corpora is English with tags of the syntactic tree form of Penn Tree Bank[7], and is obtained by parsing on Apple Pie Parser. The corpus data is stored in Postgres database server. In the trial production of the system, as experimental data, we got the English abstract currently opened to public as the web page of IEICE[8] and JAIR[9], obtained number is about 4200.

Input Sentence Analysis Program

This program parses input sentence by Apple Pie Parser like corpus data. When an input sentence includes Japanese words, this program analyzes the words by using chasen and gets synonyms and English translation words. This program gets synonyms of English by wordnet. These related words are presented when a learner demands.

Example Retrieval Program

This program searches corpus for the phrase structures that are similar with the phrase structures of the input sentence. Retrieval keyword is constituted of nouns, verbs and part-of-speech information added by the parser, and is not constituted of the words except nouns and verbs. For example, when the noun phrase '(NPL (DT the) (NN method))' is contained in the input sentence, the noun phrase '(NPL (DT a) (JJ new) (NN method))' matches it.

Example Analysis Program

This prototype system regards the last phrase of an input sentence as a support target phrase. For all the example sentences which matched the last phrase, this program gets the matched phrase and the following phrase. This program classifies it by structure of those phrases and sorts it in numerical order.

User Interface

Figure 8, 9, 10 shows a screen shot of the prototype system. This window has four tab panels that is at any time selectable. Figure 8 indicates a screen shot at having selected example sentences tab. Example sentences are displayed in field of the above and we can read the whole sentence of abstract including the selected example sentence in field of the below. The left number of an example sentence is a parameter to show a similar degree with an input sentence. The above window of figure 9 indicates a screen shot at having selected phrase structure tab and has the check box that can control a condition of an example sentence retrieval. In addition, we can revise error

of parsed result. The bottom window of figure 9 indicates a screen shot at having selected related words tab. Figure 10 indicates a screen shot at having selected following candidate phrases tab. When a sentence is amended and a renewal button is pushed, information is renewed.

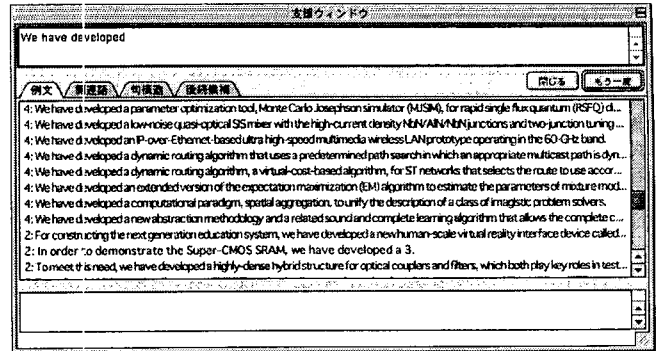


Figure 8 - Screen Shot (Example Sentences)

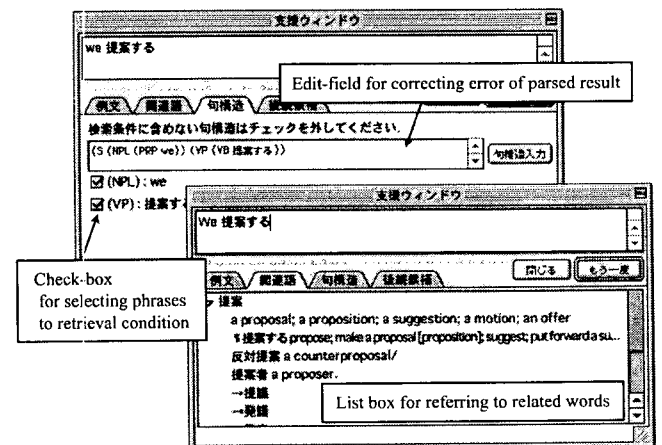


Figure 9 - Screen Shot (Phrase Structure, Related Words)

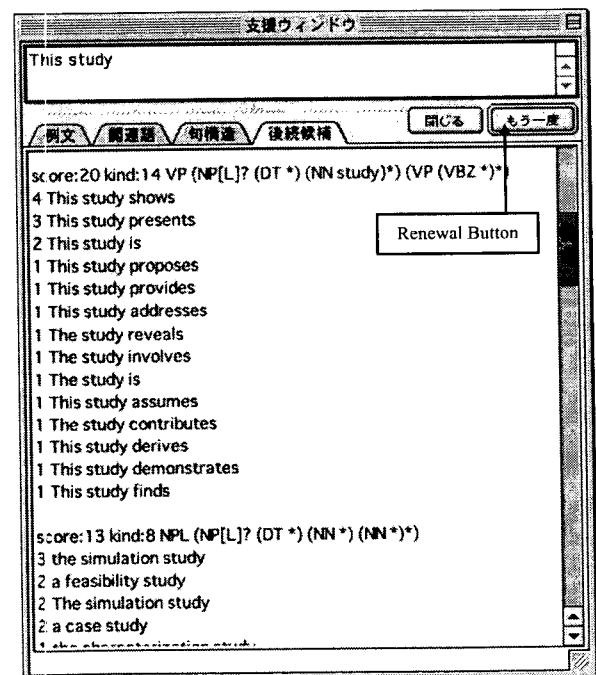


Figure 10 - Screen Shot (Following Candidate Phrases)

Experimental Result

Table 1 shows the experimental result of statistics of retrieval phrases. Input sentence is only a simple noun phrase of a determiner and a noun. A number in a parenthesis expresses number of a pattern at having classified a phrase by form of structure. As for some words, an example more than 500 cases of the upper limit was found.

Table 2 indicates statistics only of the following phrase. A number in a parenthesis expresses number of a pattern as well as Table 1. A number in a parenthesis of a row of a verb phrase expresses kind of a verb of the verb phrase.

These results represent meanings and characteristics that each words have. And enough data are provided to select the next phrase. But a learner is not confused because there is not too much number of a pattern of a candidate phrase.

Table 1 - Statistics of retrieved phrases

Noun Word	Following Phrase	Phrase Varieties	Available Example	Retrieved Example
method	110 (35)	299 (44)	409	500
function	72 (31)	315 (53)	387	493
system	54 (18)	305 (66)	359	500
paper	245 (14)	20 (6)	265	500
environment	25 (22)	183 (46)	208	282
result	126 (30)	58 (16)	184	332
mechanism	33 (22)	124 (34)	157	201
language	17 (9)	48 (20)	65	100
experiment	19 (12)	35 (14)	54	74
learner	9 (4)	5 (4)	13	14

Table 2 - Kind of the following phrase

Word	Verb	Prep	Others	Total
method	54 (10)	43 (16)	13 (9)	110 (35)
function	6 (3)	57 (21)	9 (7)	72 (31)
system	36 (4)	9 (8)	9 (6)	54 (18)
paper	222 (6)	0 (0)	23 (8)	245 (14)
environment	10 (10)	8 (7)	7 (5)	25 (22)
result	59 (6)	59 (19)	8 (5)	126 (30)
mechanism	7 (3)	17 (11)	9 (8)	33 (22)
language	11 (4)	4 (4)	2 (1)	17 (9)
experiment	9 (4)	8 (6)	2 (2)	19 (12)
learner	6 (2)	1 (1)	2 (1)	9 (4)

Conclusion

In the prototype system, we have succeeded in providing a learner with intelligible information by using a phrase structure tag similar example retrieval. A learner can train efficiently the writing skill which is difficult to acquire in a short period.

However, the reported precision about Apple Pie Parser is too low to practical use. Moreover, this parser is not suitable for analyzing input sentence including errors. Therefore, as for future work, we will develop analysis device for a sentence including errors. And we will evaluate its effectiveness.

Acknowledgments

This work was supported by a Grant-in Aid for Scientific Research (B) No.13480047 from the Ministry of Education, Science Sports and Culture in Japan.

References

- [1] Yamamoto, H.; and Kitamura, M. 1999. Corpus based natural language processing and an education system using it. *Transactions of Japanese Society for Information and Systems in Education* 16(1):43-50.
- [2] Takeda, A.; and Furugori, T. 1994. A Sample-Based System for Helping Japanese Write English Sentences. *Transactions of Information Processing Society of Japan* 35(1):53-61.
- [3] Matsumoto, Y. 1997. CHASEN: Japanese Morpheme Analysis System. <http://cactus.aist-nara.ac.jp/>
- [4] Fellbaum, C. 1998. WordNet: An Electronic Lexical Database. <http://www.cogsci.princeton.edu/~wn/>
- [5] Sekine, S.; and Grishman, R. 1997. Apple Pie Parser. <http://www.cs.nyu.edu/cs/projects/proteus/app/>
- [6] IPA Lexicon of the Japanese Language for computers. <http://www.ipa.go.jp/STC/NIHONGO/IPAL/ipal.html>
- [7] Marcus, M.; Kim, G.; Marcinkiewicz, M. A.; MacIntyre, R.; Bies, A.; Ferguson, M.; Katz, K.; and Schasberger, B. 1994. Penn TREEBANK. <http://www ldc.upenn.edu/doc/treebank2/treebank2.index.html>
- [8] The Institute of Electronics, Information and Communication Engineers, <http://www.ieice.org/eng/>
- [9] Journal of Artificial Intelligence Research, <http://www.jair.org/>
- [10] Tokuda, N.; and Chen, L. 2001. An Online Tutoring System for Language Translation. *IEEE MultiMedia* 8(3):46-55.