

A Web-based English Abstract Writing Tool Using a Tagged E-J Parallel Corpus

Masumi Narita[†] Kazuya Kurokawa[‡] Takehito Utsuro[‡]

[†]Software Research Center
Ricoh Co., Ltd.
1-1-17 Koishikawa, Bunkyo-ku
Tokyo, Japan
narita@src.ricoh.co.jp

[‡]Computational Linguistics Laboratory
Department of Information and Computer Sciences
Toyohashi University of Technology
1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Japan
kazuya@cl.ics.tut.ac.jp, utsuro@ics.tut.ac.jp

Abstract

In this paper, we present a Web-based English abstract writing tool, the “BEAR (Building English Abstracts by Ricoh).” This English writing tool is aimed at helping Japanese software engineers improve the organization of their writing by enabling them to select a rhetorical template of the target abstract and to build up component sentences while having access to good-quality sample sentences. To provide this kind of language assistance, we constructed an E-J parallel corpus of 539 sample abstracts as the core language resource. After analyzing the rhetorical structure of these sample abstracts, we tagged the corpus with textual and linguistic information. The “BEAR” is not designed for beginners but for intermediate to advanced EFL learners who very often need to write a research paper or a technical report in English. Software development has not yet been completed, but we have already gathered some user feedback at preliminary user trials. We show that the “BEAR” has been positively evaluated by our users and thus our tagged E-J parallel corpus of sample abstracts can support our users in the difficult task of working with a foreign language. We also discuss the outlook for further development of the “BEAR.”

1. Introduction

In recent studies (Narita, 1998, 2000a, 2000b), we have shown that corpus-based English writing tools have a very high potential to assist Japanese EFL learners in writing in English. Central to our argument is the assumption that frequent access to good-quality sample sentences/texts helps our users increase and improve their English proficiency autonomously in terms of sentence construction and paragraph development as well.

Under this assumption, a UNIX-based English abstract writing tool was developed using a tagged E-J parallel corpus of 539 sample abstracts and was fairly positively evaluated by our trial users (Narita, 2000b). This UNIX-based writing tool was designed to solve the main problems that foreign readers tend to identify in English documents produced by Japanese authors such as poor organization, unclear logic, unclear focus, too-lengthy paragraphs, and poorly constructed sentences (Leggett, 1966; Tomiyama and Tomiyama, 1996). Paper abstracts were selected as the target document for our tool because they need to be written in a concise, logical and coherent sequence.

Our approach to computer-assisted foreign language production was distinctly different from conventional ones in that the primary focus is to raise our users’ awareness of the rhetorical structure of English abstracts. Thus, our users were encouraged to select a specific type of organization according to their needs by quickly scanning the sample abstracts we collected before they started off the very first sentence.

As shown in Narita (2000b), the effectiveness of searching for a good model of the target writing was supported by user feedback we gathered. Nevertheless, our trial users gave us some comments and requests for the refinement of our tool. They preferred to be given more ‘user-friendly’ assistance by a Windows-based version of the tool and to be given a lexical look-up function above all other things. What is meant by more ‘user-friendly’ assistance is to make it easier for our users to decide how to structure their target abstract through their scanning of sample abstracts.

To provide more writing support for our users, we have developed a Web-based version of our English abstract writing tool, the “BEAR (Building English Abstracts by Ricoh).” This is a program designed to improve overall usability and ensure flexibility in operation, thus helping users become more productive in EFL writing. Major refinements are (1) the inclusion of rhetorical templates that guide users to more efficient and effective abstract writing, (2) the extension of the “Sample Sentence Search” function that makes possible a KWIC (Keyword in Context) search, and (3) links to some Web sites that provide on-line lexical look-up or spelling check tools.

The “BEAR” is not designed for beginners but for intermediate to advanced EFL learners who very often need to write a research paper or a technical report in English. Software development has not yet been completed, but we have already tested parts of the program with 27 trial users.

The rest of the paper is organized as follows. In Section 2, we will give a brief overview of the “BEAR.” In Section 3, we will look at the software in more detail, starting with a description of the core language resource, our tagged E-J parallel corpus of sample abstracts, and then dealing with the different search functions of the tool. In Section 4, we will present some of the user feedback we have gathered at preliminary user trials. Finally, in Section 5, we discuss our outlook for further development of the “BEAR.”

2. Overview of the “BEAR”

The “BEAR” is designed to assist Japanese software engineers who are intermediate to advanced EFL learners in writing English abstracts for their research papers or technical reports. The target domain for abstract writing is information engineering. The following features are characteristic of this tool:

- Web-based (currently accessible on the Ricoh’s Intranet)
- Template- and example-based writing prompts

- Aimed at promoting the acquisition of procedural skills in English writing
- Linked to other language resource sites

The program consists of four modules listed below.

- (1) Rhetorical Template Selection
- (2) Component Sentence Construction
 - The following search functions are used:
 - Sample Sentence Search by Sentence Roles
 - Sample Sentence Search by Keywords
 - On-Line Lexical Look-Up/Spelling Check
 - Sentence Pattern Search*
- (3) Feedback Message Generation*
- (4) Sentence Concatenation (Output Formatting)
 - *Under construction

It is important to note that all the modules except the last one rely on our tagged E-J parallel corpus of sample abstracts.

The “BEAR” works as follows. First, on the screen menu (Fig. 1), users are prompted to specify their research field, an abstract type, and the location of the topic sentence in the target abstract. Guidance information is available on demand.

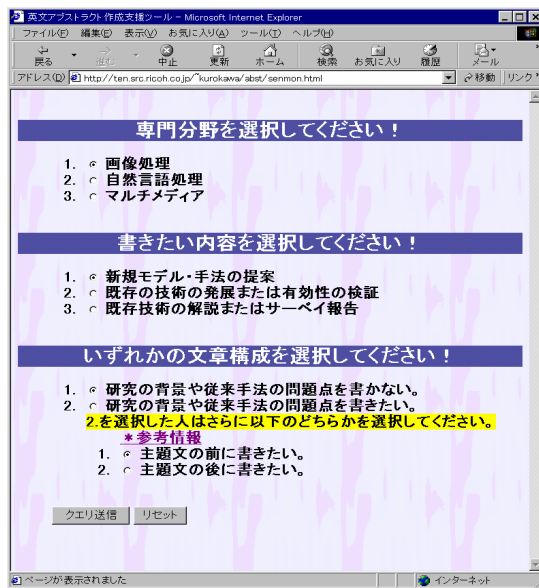


Fig. 1: Screen Menu for Rhetorical Template Selection

When all the information is specified, a rhetorical template is shown on the screen (Fig. 2) that includes several separate windows for producing component sentences of the target abstract. Since necessary components are differentiated from optional ones by the color of the window, users will be given a warning message if they fail to complete the necessary ones.

The next step to be taken is to try to produce each component sentence by freely using our “Sample Sentence Search,” “On-Line Lexical Look-Up/Spelling Check,” and “Sentence Pattern Search” (under construction) functions. Two kinds of “Sample Sentence Search” are possible with the “BEAR.” One is driven by a sentence role and the other by a keyword in English or in Japanese.



Fig. 2: A Rhetorical Template Retrieved

On-line lexical look-up or spelling check service sites are listed in the left frame of the screen. Users can jump to the target site by clicking on a list item. The “Sentence Pattern Search,” which is under construction, is designed not only to give information on possible complementation patterns of a given verb but also to retrieve sample sentences of a specified complementation pattern from our corpus.

Users are encouraged to find a good model of their target sentence by scanning each of the sample sentences retrieved from a subset of our corpus (due to the user input on the menu in Fig. 1) by our search engines and start writing by ‘borrowing,’ that is, modifying the sample sentence that they have chosen as a good model so as to express their own ideas. When all the component sentences are produced, they are concatenated as an abstract with a click of the [FINISH] button.

The “Feedback Message Generation” module, which is also under construction, will be discussed later in Section 5.

3. Language Resources and Search Functions of the “BEAR”

Since the “BEAR” is aimed at helping our users search for a good model of their target abstract, target sentence or target expression, our corpus building of well-organized and good-quality sample English abstracts is the key to success.

The subsequent subsections describe how we built our tagged E-J parallel corpus of sample abstracts as the core language resource and how our different search functions are running or expected to run on the “BEAR.”

3.1. Building a Tagged E-J Parallel Corpus of Sample Abstracts

With permission to use them for research purposes, we collected a total of 539 sample English abstracts from widely known technical journals (“*IEEE Transactions on Pattern Analysis and Machine Intelligence*” and “*IEEE Multimedia*”) and conference proceedings (“*Proceedings of the Annual Meeting of ACL*”).

Japanese translations of sample English abstracts were prepared to make it easier for users to search for a good model of the target sentence or the target expression. These Japanese equivalents were voluntarily constructed on a sentence-to-sentence basis by Ricoh's software engineers, thereby aligning English-Japanese sentence pairs of sample abstracts manually. Since the engineers are well informed about the topic areas, they could produce high-quality Japanese equivalents.

We then examined how to tag our corpus by analyzing the sample abstracts we collected in terms of their textual structure and logical sequence as well as their writing quality.

Thus, we designed our corpus to be marked up and tagged with the following information in an SGML-conformant way:

- (1) Text Features --- Internal Organization
- (2) Bibliographic Information
- (3) Writing Quality (High/Mid/Low)
- (4) Linguistic Information
 - (4-1) Abstract Types
 - (4-2) Organizational-Scheme Types
 - (4-3) Sentence Roles
 - (4-4) Verb Complementation Patterns

Each sample abstract was linguistically tagged with its abstract type and its organizational-scheme type. Abstract types represent what the authors intend to convey in their papers and organizational-scheme types represent the location of the topic sentence in the abstract.

Each sample sentence was tagged with its sentence role and verb complementation pattern(s). Sentence roles were assigned according to the logical relationships with the topic sentence. Verb complementation patterns were assigned based on the COMLEX Syntax V2.2, a computational lexicon that was developed by Grishman *et al.* (1994) at New York University.

Writing quality of the sample abstracts was evaluated by an expert proofreader of research papers, a native speaker of English. Low-quality samples were used for our linguistic analysis only to design our "Feedback Message Generation" module.

When this tagging was completed, we extracted only the information on verb complementation patterns to build a separate lexical database with their frequency counts in our corpus. This lexical database is scheduled to be linked to our corpus of sample abstracts so that sample sentences of a specified complementation pattern can be retrieved from our corpus at the user's request.

A detailed explanation of our tagsets and our sample tagging are referred to Narita (2000a, 2000b).

3.2. Sample Sentence Search Functions

The "BEAR" is designed to help users find a good model of their target sentence by providing two kinds of "Sample Sentence Search" functions: (1) the "Sample Sentence Search by Sentence Roles" function and (2) the "Sample Sentence Search by Keywords" function.

Sentence role-driven search button is located on the right of the description of the sentence role given to each window for sentence construction. In Fig. 2, for instance, clicking on the search button of the top window opens a new screen, where the first sample sentence, whose

sentence role is the topic sentence, is retrieved from our E-J parallel corpus and shown in red, one at a time, with adjacent sentences in black (Fig. 3). Users are encouraged to scan other sample sentences retrieved until they end up with a good model to 'borrow.'

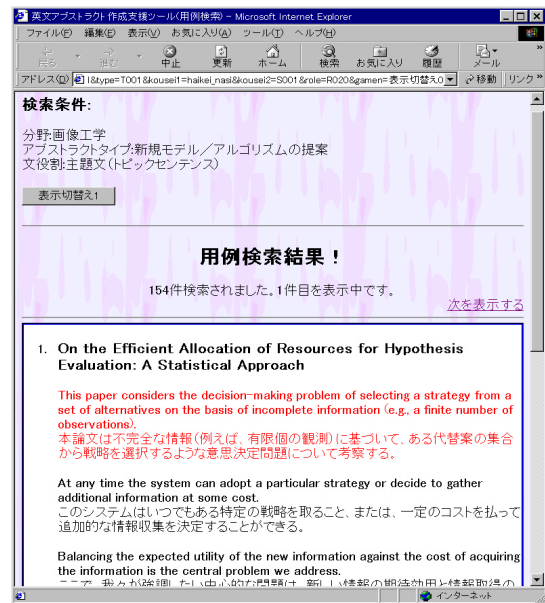


Fig. 3: Sample Sentence Search by Sentence Roles

Keyword-driven search button is located on the upper right corner of the screen in Fig. 2. When clicking on this search button, users are prompted to input a keyword in English or in Japanese. With the input of a keyword, sample sentences from our E-J corpus are shown in a 'concordanced' (KWIC) form (Fig. 4). The sample sentence can be accessed within the whole abstract by clicking on a specific line on the concordance list.

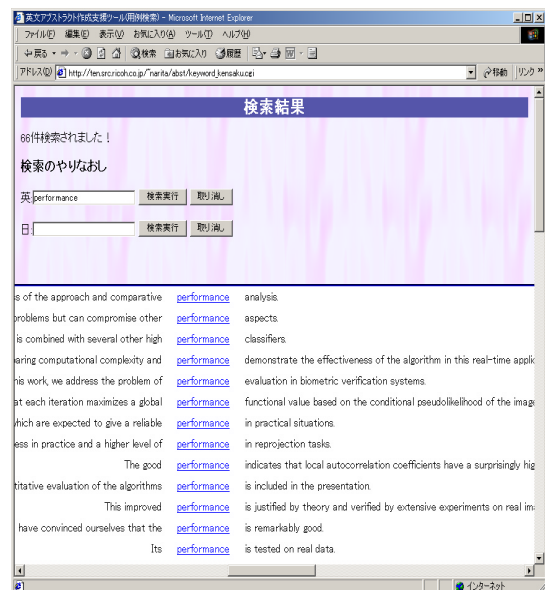


Fig. 4: Sample Sentence Search by Keywords

Concordance lists of this kind can also be used for finding collocational patterns that a given keyword is likely to call for.

3.3. Sentence Pattern Search Function

As mentioned in 3.1, we built a separate lexical database of verb complementation patterns from our E-J parallel corpus of sample abstracts. An example of this tagging in our corpus is shown in bold-face below:

These features [**make@@NP1 make NP2 NP3 PRED@@**] the algorithm a useful tool for the quantitative analysis of real-world images.

In this lexical database, each verb entry is given a list of possible complementation patterns with their frequency counts in our corpus. The corpus size, however, is rather small, so that we plan to include the language resources developed by Narita (1998).

When verb complementation patterns are automatically linked to their respective sample sentences in our corpus, users can easily retrieve the sample sentences by specifying the pattern they need.

4. User Feedback

We evaluated the usability of the “BEAR” in spite of the fact that some components are still under construction. We asked 27 software engineers at Ricoh to use the “BEAR” for about an hour.

User feedback was obtained by two means: (1) questionnaire measure of user satisfaction and users’ perceived utility of software components, and (2) asking users to write their reactions to our tool in their own words. We designed a questionnaire in which several factors were included that may affect computer user satisfaction. Our trial users were asked to rate their satisfaction and their perceived utility on a five-point scale (-2, -1, 0, +1, +2) so that an individual’s feeling can be placed somewhere between a “most positive” reaction and a “most negative” reaction.

Tables 1 and 2 show the average rating for each factor and that for each software component, respectively.

Factor	Average Rating
Effectiveness of the Tool	+1.19
Response Time	+1.11
Effectiveness of Samples	+0.93
Ease of Use	+0.26
Display Format	+0.15
Volume of Samples	-0.30

Table 1: User Satisfaction with the “BEAR”

Software Component	Average Rating
Rhetorical Template Selection	+0.70
Sample Sentence Search (Sentence Role-Driven)	+0.78
Sample Sentence Search (Keyword-Driven)	+1.26
Lexical Look-Up/Spelling Check	+1.29

Table 2: Users’ Perceived Utility of Software Component

The factors “Effectiveness of the Tool” and “Response Time” were positively evaluated by our trial users, whereas it turned out that we need to improve the factors “Ease of Use” and “Display Format.” Our oral interviews with trial users also reinforced the necessity to enhance the appearance and effectiveness of our user interface. The factor “Volume of Samples” was negatively evaluated, which clearly means that we need to provide a larger number of samples.

Among our software components, the “Lexical Look-Up/Spelling Check” and the “Sample Sentence Search by Keywords” were fairly positively evaluated. Apparently, our trial users preferred to access the sample sentences via the keyword input. This suggests that we should design the “Sample Sentence Search” function in such a way that the keyword-driven search can be freely combined with the sentence role-driven search.

The “Rhetorical Template Selection” was less positively evaluated than we expected. However, some of our trial users, who had used our UNIX-based writing tool, pointed out that the template eased their responsibility of deciding the organization of their target abstract on their own.

Our trial users also gave us the following comments and requests in their own words:

- They need much easier access to their relevant sample sentences.
- They need a more sophisticated user interface.
- They need to access a larger number of samples within a specific domain.
- They need to access corpora from a wider range of information science domains.
- They need computer support for grammar checking.
- They need computer support for transition markers.
- They need to access most commonly occurring errors in English texts produced by Japanese authors.
- They prefer to have a “fill-in-the-blank”-fashioned rhetorical template.

5. Conclusion and Future Work

We developed a prototype of the “BEAR,” which is currently accessible via Intranet to all the software engineers at R&D laboratories of Ricoh.

The “BEAR” provides users with relevant information to help them produce a well-organized abstract, as well as well-formed English component sentences in the abstract. To give our users both discourse-level and sentence-level assistance in an organized way, our tagged E-J parallel corpus of sample abstracts plays an essential role as the core language resource for this writing tool.

Software development has not yet been completed, but we have already tested parts of the program with 27 software engineers at Ricoh. Our preliminary user trials show that the “BEAR” has been positively evaluated by our users and thus our tagged E-J parallel corpus of sample abstracts can support our users in the difficult task of working with a foreign language.

There is clearly further to go. First, we need to make available the “Sentence Pattern Search” function and the “Feedback Message Generation” module. In particular, the “Feedback Message Generation” module should be carefully designed to promote users’ awareness of grammar pitfalls that they are likely to fall into when

writing in English. A number of on-line grammar handbooks or language resources are available on the Web, but what we aim to do is to interactively help users acquire how to express their own ideas in a more effective way while avoiding rhetorical or grammatical deviations that might cause misunderstanding.

Quite recently, in SLA (Second Language Acquisition) studies, there was a heated debate between Truscott (1999) and Ferris (1999) on the effectiveness of grammar feedback. Along with this debate, many questions remain whether grammar feedback should be abandoned, whether only the number or the location of errors should be given to EFL learners and so forth. Since our oral interviews with trial users have revealed that our users tend to spend a lot of time trying to avoid grammatical errors rather than trying to concentrate on their ideas, the "BEAR" is required to become an efficient grammar tutor.

Given a very robust parser, for instance, users' erroneous sentences or sentence fragments are parsed to identify the source of grammatical errors. If this kind of information is stored as a user profile for a given period of time, our "Feedback Message Generation" module can possibly be personalized.

Second, we need to look at how our users make use of the "BEAR" over a period of time, what they perceive as their long-term benefits and whether these lead to more successful abstract writing. Logging how our users interact with the "BEAR" may well promote our understanding of how their procedural skills develop and how the "BEAR" can best help.

Third, we plan to work on redesigning the "BEAR" so that users can begin with a rough plan and freely revise and extend the plan later as the "Sketcher" defined by Wyllie (2000). In other words, we plan to provide a writing tool that can support users throughout their writing process (Raimes, 1992) because writing is a complex task involving 'juggling with simultaneous constraints' in the words of Flower and Hayes (1980). In this respect, the rationale behind the redesign of the "BEAR" is similar to that of the "Academic Writer" developed at University of Brighton (Broady and Shurville, 2000).

Further, we need to make a practical and precise study of a wider range of genres in information engineering, in particular, in terms of the rhetorical structure and grammar conventions.

Finally, in order to efficiently broaden the coverage of our language resources, we will be developing the possibility of semi-automated corpus tagging, based on our experiences in manual tagging.

Acknowledgements

This work was partly supported by Communications Research Laboratory in Japan. We would like to thank Dr. Hitoshi Isahara for his valuable contributions and support.

References

Broady, E. and Shurville, S. (2000). Developing Academic Writer: Designing a writing environment for novice academic writers. In E. Broady (Ed.) *Second Language Writing in a Computer Environment* (pp. 117--137). Centre for Information on Language Teaching and Research.

- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing* (pp. 1--11).
- Flower, L. S. and Hayes, J. R. (1980). The Dynamics of Composing: Making Plans and Juggling Constraints. In L. W. Gregg and E. R. Steinberg (Eds.) *Cognitive Processes in Writing* (pp. 31--50). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Grishman, R., Macleod, C. and Mayers, A. (1994). Complex Syntax: Building Computational Lexicon. In *Proceedings of COLING-94* (pp. 268--272).
- Leggett, A. J. (1966). Notes on the Writing of Scientific English for Japanese Physicists. *Journal of the Physical Society of Japan* (pp. 790--805).
- Narita, M. (1998). Language Resources for Writer's Helper. In *Proceedings of the 1st International Conference on Language Resources and Evaluation* (pp. 269--273).
- Narita, M. (2000a). Constructing a Tagged E-J Parallel Corpus for Assisting Japanese Software Engineers in Writing English Abstracts. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation* (pp. 1187--1191).
- Narita, M. (2000b). A Corpus-based English Language Assistant to Japanese Software Engineers. In *Proceedings of MT-2000 Conference* (pp. 24-1--24-8).
- Raimes, A. (1992). *Exploring Through Writing: a Process Approach to ESL Composition*. New York: St. Martins Press.
- Tomiyama M. and Tomiyama K. (1996). *Rikoukei Eigo Ronbun to Koutouhappyou no Jissai*. Tokyo: Corona Publishing Co., Ltd.
- Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes.": A response to Ferris. *Journal of Second Language Learning* (pp. 111--122).
- Wyllie, A. (2000). 'On the Road to Discovery': A Study of the Composing Strategies of Native and Non-Native Academic Writers Using the Word Processor. In E. Broady (Ed.) *Second Language Writing in a Computer Environment* (pp. 95--116). Centre for Information on Language Teaching and Research.