Text Simplification for Reading Assistance: A Project Note

Kentaro Inui Atsushi Fujita Tetsuro Takahashi Ryu Iida
Nara Advanced Institute of Science and Technology
Takayama, Ikoma, Nara, 630-0192, Japan
{inui,atsush-f,tetsu-ta,ryu-i}@is.aist-nara.ac.jp

Tomoya Iwakura
Fujitsu Laboratories Ltd.
Kamikodanaka, Nakahara, Kawasaki, Kanagawa, 211-8588, Japan
iwakura.tomoya@jp.fujitsu.com

Abstract

This paper describes our ongoing research project on text simplification for congenitally deaf people. Text simplification we are aiming at is the task of offering a deaf reader a syntactic and lexical paraphrase of a given text for assisting her/him to understand what it means. In this paper, we discuss the issues we should address to realize text simplification and report on the present results in three different aspects of this task: readability assessment, paraphrase representation and post-transfer error detection.

1 Introduction

This paper reports on our ongoing research into text simplification for reading assistance. Potential users targeted in this research are congenitally deaf people (more specifically, students at (junior-)high schools for the deaf), who tend to have difficulties in reading and writing text. We are aiming at the development of the technology of text simplification with which a reading assistance system lexically and structurally paraphrases a given text into a simpler and plainer one that is thus more comprehensible.

The idea of using paraphrases for reading assistance is not necessarily novel. For example, Carroll *et al.* (1998) and Canning and Taito (1999) report on their project in which they address syntactic transforms aiming at making newspaper text accessible to aphasics. Following this trend of research, in this project, we address four unexplored issues as below besides the user- and task-oriented evaluation of the overall system.

Before going to the detail, we first clarify the four issues we have addressed in the next section. We then reported on the present results on three of the four, readability assessment, paraphrase representation and post-transfer error detection, in the subsequent sections.

2 Research issues and our approach

2.1 Readability assessment

The process of text simplification for reading assistance can be decomposed into the following three subprocesses:

a. *Problem identification*: identify which portions of a given text will be difficult for a given user to read,
b. *Paraphrase generation*: generate possible candidate paraphrases from the identified portions, and
c. *Evaluation*: re-assess the resultant texts to choose the one in which the problems have been resolved.

Given this decomposition, it is clear that one of the key issues in reading assistance is the problem of assessing the readability or comprehensibility[1] of text because it is involved in subprocesses (a) and (c).

Readability assessment is doubtlessly a tough issue (Williams *et al.*, 2003). In this project, however, we argue that, if one targets only a particular population segment and if an adequate collection of data is available, then corpus-based empirical approaches may well be feasible. We have already proven that one can collect such readability assessment data by conducting survey questionnaires targeting teachers at schools for the deaf.

[1] In this paper, we use the terms *readability* and *comprehensibility* interchangeably, while strictly distinguishing them from *legibility* of each fragment (typically, a sentence or paragraph) of a given text.

2.2 Paraphrase acquisition

One of the good findings that we obtained through the aforementioned surveys is that there are a broad range of paraphrases that can improve the readability of text. A reading assistance system is, therefore, hoped to be able to generate sufficient varieties of paraphrases of a given input. To create such a system, one needs to feed it with a large collection of paraphrase patterns. Very timely, the acquisition of paraphrase patterns has been actively studied in recent years:

- Manual collection of paraphrases in the context of language generation, e.g. (Robin and McKeown, 1996),
- Derivation of paraphrases through existing lexical resources, e.g. (Kurohashi *et al.*, 1999),
- Corpus-based statistical methods inspired by the work on information extraction, e.g. (Jacquemin, 1999; Lin and Pantel, 2001), and
- Alignment-based acquisition of paraphrases from comparable corpora, e.g. (Barzilay and McKeown, 2001; Shinyama *et al.*, 2002; Barzilay and Lee, 2003).

One remaining issue is how effectively these methods contribute to the generation of paraphrases in our application-oriented context.

2.3 Paraphrase representation

One of the findings obtained in the previous studies for paraphrase acquisition is that the automatic acquisition of candidates of paraphrases is quite realizable for various types of source data but acquired collections tend to be rather noisy and need manual cleaning as reported in, for example, (Lin and Pantel, 2001). Given that, it turns out to be important to devise an effective way of facilitating manual correction and a standardized scheme for representing and storing paraphrase patterns as shared resources.

Our approach is (a) to define first a fully expressible formalism for representing paraphrases at the level of tree-to-tree transformation and (b) devise an additional layer of representation on its top that is designed to facilitate handcoding transformation rules.

2.4 Post-transfer text revision

In paraphrasing, the morpho-syntactic information of a source sentence should be accessible throughout the transfer process since a morpho-syntactic transformation in itself can often be a motivation or goal of paraphrasing. Therefore, such an approach as semantic transfer, where morpho-syntactic information is highly abstracted away as in (Dorna *et al.*, 1998; Richardson *et al.*, 2001), does not suit this task. Provided that the morpho-syntactic stratum be an optimal level of abstraction for representing paraphrasing/transfer patterns, one must recall that semantic-transfer approaches such as those cited above were motivated mainly by the need for reducing the complexity of transfer knowledge, which could be unmanageable in morpho-syntactic transfer.

Our approach to this problem is to (a) leave the description of each transfer pattern *underspecified* and (b) implement the knowledge about linguistic constraints that are independent of a particular transfer pattern separately from the transfer knowledge. There are a wide range of such transfer-independent linguistic constraints. Constraints on morpheme connectivity, verb conjugation, word collocation, and tense and aspect forms in relative clauses are typical examples of such constraints.

These four issues can be considered as different aspects of the overall question how one can make the development and maintenance of a gigantic resource for paraphrasing tractable. (1) The introduction of readability assessment would free us from cares about the purposiveness of each paraphrasing rule in paraphrase acquisition. (2) Paraphrase acquisition is obviously indispensable for scaling up the resource. (3) A good formalism for representing paraphrasing rules would facilitate the manual refinement and maintenance of them. (4) Post-transfer error detection and revision would make the system tolerant to flows in paraphrasing rules.

While many researchers have addressed the issue of paraphrase acquisition reporting promising results as cited above, the other three issues have been left relatively unexplored in spite of their significance in the above sense. Motivated by this context, in the rest of this paper, we address these remaining three.

3 Readability assessment

To the best of our knowledge, there have never been no reports on research to build a computational model of the language proficiency of deaf people, except for the remarkable reports by Michaud and McCoy (2001). As a subpart of their research aimed at developing the ICICLE system (McCoy and Masterman, 1997), a language-tutoring application for deaf learners of written English, Michaud and McCoy developed an architecture for modeling the writing proficiency of a user called SLALOM. SLALOM is designed to capture the stereotypic linear order of acquisition within certain categories of morphological and/or syntactic features of language. Unfortunately, the modeling method used in SLALOM cannot be directly applied to our domain for three reasons.

- Unlike writing tutoring, in reading assistance, target sentences are in principle unlimited. We therefore need to take a wider range of morpho-syntactic features into account.
- SLALOM is not designed to capture the difficulty of any combination of morpho-syntactic features, which it is essential to take into account in reading assistance.
- Given the need to consider feature combinations, a simple linear order model that is assumed in SLALOM is unsuitable.

3.1 Our approach: We ask teachers

To overcome these deficiencies, we took yet another approach where we designed a survey questionnaire targeting teachers at schools for the deaf, and have been collecting readability assessment data. In this questionnaire, we ask the teachers to compare the readability of a given sentence with *paraphrases* of it. The use of paraphrases is of critical importance in our questionnaire since it makes manual readability assessment significantly easier and more reliable.

3.1.1 Targets

We targeted teachers of Japanese or English literacy at schools for the deaf for the following reasons.

Ideally, this sort of survey would be carried out by targeting the population segment in question, i.e., deaf students in our study. In fact, pedagogists and psycholinguists have made tremendous efforts to examine the language proficiency of deaf students by giving them proficiency tests. Such efforts are very important, but they have had difficulty in capturing enough of the picture to develop a comprehensive and implementable reading proficiency model of the population due to the expense of extensive language proficiency testing.

In contrast, our approach is an attempt to model the knowledge of experts in this field (i.e., teaching deaf students). The targeted teachers have not only rich experiential knowledge about the language proficiency of their students but are also highly skilled in paraphrasing to help their students' comprehension. Since such knowledge gleaned from individual experiences already has some generality, extracting it through a survey should be less costly and thus more comprehensive than investigation based on language proficiency testing.

3.1.2 Questionnaire

In the questionnaire, each question consists of several paraphrases, as shown in Figure 1 (a), where (A) is a source sentence, and (B) and (C) are paraphrases of (A). Each respondent was asked to assess the *relative* readability of the paraphrases given for each source sentence, as shown in Figure 1 (b). The respondent judged sentence (A) to be the most difficult and judged (B) and (C) to be comparable. A judgment that sentence s_i is easier than sentence s_j means that s_i is judged likely to be understood by a larger subset of students than s_j. We asked the respondents to annotate the paraphrases with format-free comments, giving the reasons for their judgments, alternative paraphrases, etc., as shown in Figure 1 (b).

To make our questionnaire efficient for model acquisition, we had to carefully control the variation in paraphrases. To do that, we first selected around 50 morpho-syntactic features that are considered influential in sentence readability for deaf people. For each of those features, we collected several simple example sentences from various sources (literacy textbooks, grammar references, etc.). We then manually produced several paraphrases from each of the collected sentences so as to remove the feature that characterized the source sentence from each paraphrase. For example, in Figure 1, the feature characterizing sentence (A) is a non-restrictive relative clause (i.e., sentence (A) was selected as an example of this feature). Neither (B) nor (C) has this feature. We also controlled the lexical variety to minimize the effect of lexical factors on readability; we also restricted the vocabulary to a top-2000 basic word set (NIJL, 1991).

3.1.3 Administration

We administrated a preliminary survey targeting three teachers. Through the survey, we observed that (a) the teachers largely agreed in their assessments of relative readability, (b) their format-free comments indicated that the observed differences in readability were largely explainable in terms of the morpho-syntactic features we had prepared, and (c) a larger-scaled survey was needed to obtain a statistically reliable model. Based on these observations, we conducted a more comprehensive survey, in which we prepared 770 questions and sent questionnaires with a random set of 240 of them to teachers of Japanese or English literacy at 50 schools for the deaf. We

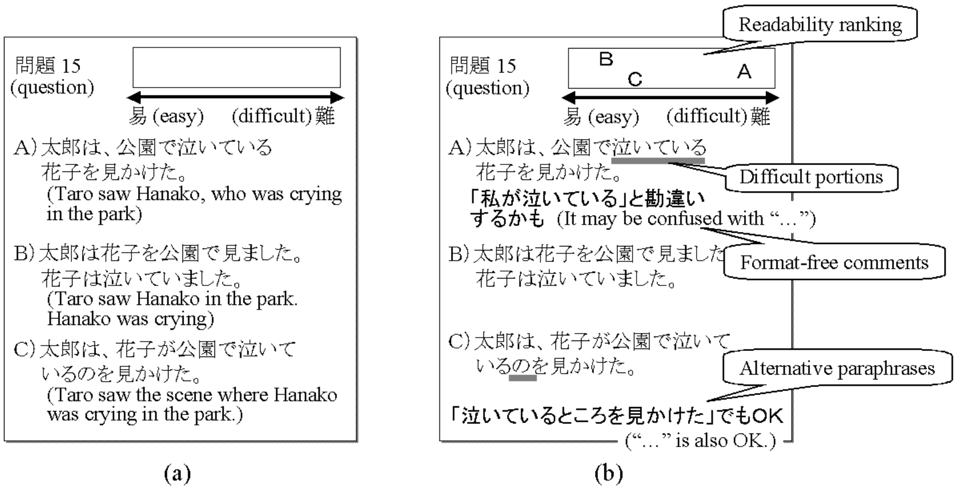

Figure 1: Sample question and response

asked them to evaluate as many as possible anonymously. We obtained 4080 responses in total (8.0 responses per question).

3.2 Readability ranking model

The task of ranking a set of paraphrases can be decomposed into comparisons between two elements combinatorially selected from the set. We consider the problem of judging which of a given pair of paraphrase sentences is more readable/comprehensible for deaf students. More specifically, given paraphrase pair (s_i, s_j), our problem is to classify it into either *left* (s_i is easier), *right* (s_j is easier), or *comparable* (s_i and s_j are comparable).

Once the problem is formulated this way, we can use various existing techniques for classifier learning. So far, we have examined a method of using the support vector machine (SVM) classification technique.

A training/testing example is paraphrase pair (s_i, s_j) coupled with its quantified class label $D(s_i, s_j) \in [-1, 1]$. Each sentence s_i is characterized by a binary feature vector F_{s_i}, and each pair (s_i, s_j) is characterized by a triple of feature vectors $\langle F^C_{s_is_j}, F^L_{s_is_j}, F^R_{s_is_j} \rangle$, where

- $F^C_{s_is_j} = F_{s_i} \wedge F_{s_j}$ (features shared by s_i and s_j),
- $F^L_{s_is_j} = F_{s_i} \wedge \overline{F_{s_j}}$ (features belonging only to s_i),
- $F^R_{s_is_j} = \overline{F_{s_i}} \wedge F_{s_j}$ (features belonging only to s_j).

$D(s_i, s_j)$ represents the difference in readability between s_i and s_j; it is computed in the following way.

1. Let $T_{s_is_j}$ be the set of respondents who assessed (s_i, s_j).
2. Given the degree of readability respondent t assigned to s_i (s_j), map it to real value $dor(t, s) \in [0, 1]$ so that the lowest degree maps to 0 and the highest degree maps to 1. For example, the degree of readability assigned to (A) in Figure 1 (b) maps to around 0.1, whereas that assigned to (B) maps to around 0.9.
3. $D(s_i, s_j) = \frac{1}{|T_{s_is_j}|} \sum_{t \in T_{s_is_j}} dor(t, s_i) - dor(t, s_j)$.

Output score $Sc_M(s_i, s_j) \in [-1, 1]$ for input (s_i, s_j) was given by the normalized distance between (s_i, s_j) and the hyperplane.

3.3 Evaluation and discussion

To evaluate the two modeling methods, we conducted a ten-fold cross validation on the set of 4055 paraphrase pairs derived from the 770 questions used in the survey. To create a feature vector space, we used 355 morpho-syntactic features. Feature annotation was done semi-automatically with the help of a morphological analyzer and dependency parser.

The task was to classify a given paraphrase pair into either *left*, *right*, or *comparable*. Model M's output class for (s_i, s_j) was given by

$$Cls_M(s_i, s_j) = \begin{cases} left & (Sc_M(s_i, s_j) \leq -\theta_m) \\ right & (Sc_M(s_i, s_j) \geq \theta_m) \\ comparable & (\text{otherwise}) \end{cases},$$

where $\theta_m \in [-1, 1]$ is a variable threshold used to balance precision with recall.

We used the 473 paraphrase pairs that satisfied the following conditions:

- $|D(s_i, s_j)|$ was not less than threshold θ_a ($\theta_a = 0.5$). The answer of (s_i, s_j) is given by
$$Cls_{Ans}(s_i, s_j) = \begin{cases} left & (D(s_i, s_j) \leq -\theta_a) \\ right & (D(s_i, s_j) \geq \theta_a) \end{cases}.$$
- (s_i, s_j) must have been assessed by more then one respondent, i.e., $|T_{s_is_j}| > 1$.

- Agreement ratio $Agr(s_i, s_j)$ must be sufficiently high, i.e., $Agr(s_i, s_j) \geq 0.9$, where $Agr(s_i, s_j) = (for(s_i, s_j) - agst(s_i, s_j))/ |T_{s_i s_j}|$, and $for(s_i, s_j)$ and $agst(s_i, s_j)$ are the number of respondents who agreed and disagreed with $Cls_{Ans}(s_i, s_j)$, respectively.

We judged output class $Cls_M(s_i, s_j)$ correct if and only if $Cls_M(s_i, s_j) = Cls_{Ans}(s_i, s_j)$. The overall performance was evaluated based on recall Rc and precision Pr:

$$Rc = \frac{\left|\{(s_i,s_j)|\ Cls_M(s_i, s_j) \text{ is correct}\}\right|}{\left|\{(s_i,s_j)|\ Cls_{Ans}(s_i,s_j) \in \{\mathit{left}, \mathit{right}\}\}\right|}$$

$$Pr = \frac{\left|\{(s_i,s_j)|\ Cls_M(s_i, s_j) \text{ is correct}\}\right|}{\left|\{(s_i,s_j)|\ Cls_M(s_i,s_j) \in \{\mathit{left}, \mathit{right}\}\}\right|}.$$

The model achieved 95% precision with 89% recall. This result confirmed that the data we collected through the questionnaires were reasonably noiseless and thus generalizable. Furthermore, both models exhibited a clear trade-off between recall and precision, indicating that their output scores can be used as a confidence measure.

4 Paraphrase representation

We represent paraphrases as transfer patterns between dependency trees. In this section, we propose a three-layered formalism for representing transfer patterns.

4.1 Types of paraphrases of concern

There are various levels of paraphrases as the following examples demonstrate:

(1) a. She *burst into tears*, and he tried to *comfort* her.
b. She *cried*, and he tried to *console* her.

(2) a. *It was* a Honda *that* John *sold to* Tom.
b. John *sold* a Honda *to* Tom.
c. Tom *bought* a Honda *from* John.

(3) a. They got married *three years ago*.
b. They got married *in 2000*.

Lexical vs. structural paraphrases Example (1) includes paraphrases of the single word "*comfort*" and the canned phrase "*burst into tears*". The sentences in (2), on the other hand, exhibit structural and thus more general patterns of paraphrasing. Both types of paraphrases, *lexical and structural paraphrases*, are considered useful for many applications including reading assistance and thus should be in the scope our discussion.

Atomic vs. compositional paraphrases The process of paraphrasing (2a) into (2c) is compositional because it can be decomposed into two subprocesses, (2a) to (2b) and (2b) to (2c). In developing a resource for paraphrasing, we have only to cover non-compositional (i.e., *atomic*) paraphrases. Compositional paraphrases can be handled if an additional computational mechanism for combining atomic paraphrases is devised.

Meaning-preserving vs. reference-preserving paraphrases It is also useful to distinguish reference-preserving paraphrases from meaning-preserving ones. The above example in (3) is of the reference-preserving type. This types of paraphrasing requires the computation of reference to objects outside discourse and thus should be excluded from our scope for the present purpose.

4.2 Dependency trees (MDSs)

Previous work on transfer-based machine translation (MT) suggests that the dependency-based representation has the advantage of facilitating syntactic transforming operations (Meyers *et al.*, 1996; Lavoie *et al.*, 2000). Following this, we adopt dependency trees as the internal representations of target texts. We suppose that a dependency tree consists of a set of nodes each of which corresponds to a lexeme or compound and a set of edges each of which represents the dependency relation between its ends. We call such a dependency tree a morpheme-based dependency structure (MDS). Each node in an MDS is supposed to be annotated with an open set of typed features that indicate morpho-syntactic and semantic information. We also assume a type hierarchy in dependency relations that consists of an open set of dependency classes including *dependency*, *compound*, *parallel*, *appositive* and *insertion*.

4.3 Three-layered representation

Previous work on transfer-based MT systems (Lavoie *et al.*, 2000; Dorna *et al.*, 1998) and alignment-based transfer knowledge acquisition (Meyers *et al.*, 1996; Richardson *et al.*, 2001) have proven that transfer knowledge can be best represented by declarative structure mapping (transforming) rules each of which typically consists of a pair of source and target partial structures as in the middle of Figure 2.

Adopting such a tree-to-tree style of representation, however, one has to address the issue of the trade-off between expressibility and comprehensibility. One may want a formalism of structural

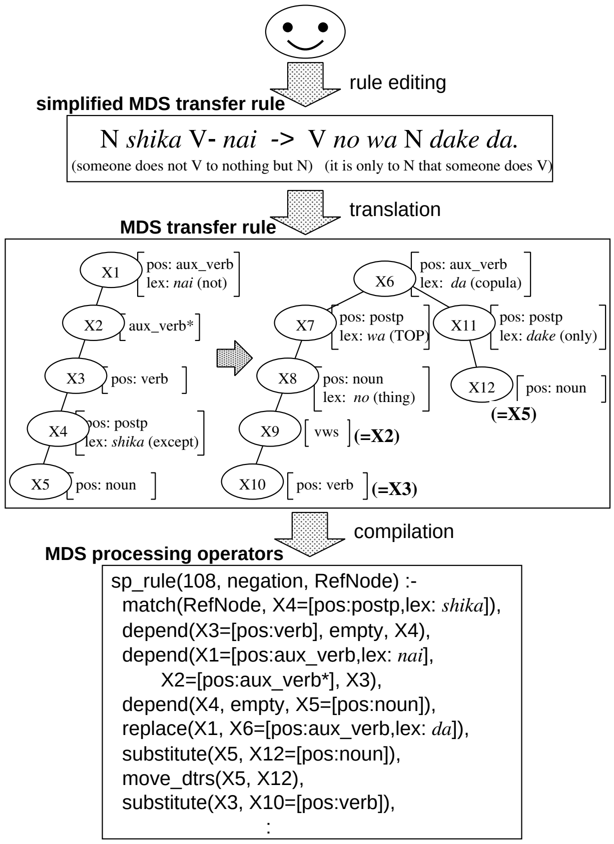

Figure 2: Three-layered rule representation

transformation patterns that is powerful enough to represent a sufficiently broad range of paraphrase patterns. However, highly expressible formalisms would make it difficult to create and maintain rules manually.

To mediate this trade-off, we devised a new layer of representation to add on the top of the layer of tree-to-tree pattern representation as illustrated in Figure 2. At this new layer, we use an extended natural language to specify transformation patterns. The language is designed to facilitate the task of hand-coding transformation rules. For example, to define the tree-to-tree transformation pattern given in the middle of Figure 2, a rule editor needs only to specify its simplified form:

(4) N *shika* V- *nai* → V *no ha* N *dake da*.
(*Someone does V to nothing but N → It is only to N that someone does V*)

A rule of this form is then automatically translated into a fully-specified tree-to-tree transformation rule. We call a rule of the latter form an MDS rewriting rule (SR rule), and a rule of the former form a simplified SR rule (SSR rule).

The idea is that most of the specifications of an SR rule can usually be abbreviated if a means to automatically complement it is provided. We use a parser and macros to do so; namely, the rule translator complements an SSR rule by macro expansion and parsing to produce the corresponding SR rule specifications. The advantages of introducing the SSR rule layer are the following:

- The SSR rule formalism allows a rule writer to edit rules with an ordinary text editor, which makes the task of rule editing much more efficient than providing her/him with a GUI-based complex tool for editing SR rules directly.
- The use of the extended natural language also has the advantage in improving the readability of rules for rule writers, which is particularly important in group work.
- To parse SSR rules, one can use the same parser as that used to parse input texts. This also improves the efficiency of rule development because it significantly reduces the burden of maintaining the consistency between the POS-tag set used for parsing input and that used for rule specifications.

The SSR rule layer shares underlying motivations with the formalism reported by Hermjakob *et al.* (2002). Our formalism is, however, considerably extended so as to be licensed by the expressibility of the SR rule representation and to be annotated with various types of rule applicability conditions including constraints on arbitrary features of nodes, structural constraints, logical specifications such as disjunction and negation, closures of dependency relations, optional constituents, etc.

The two layers for paraphrase representation are fully implemented on our paraphrasing engine KURA (Takahashi et al., 2001) coupled with another layer for processing MDSs (the bottom layer illustrated in Figure 2). The whole system of KURA and part of the transer rules implemented on it (see Section 5 below) are available at http://cl.aist-nara.ac.jp/lab/kura/doc/.

5 Post-transfer error detection

What kinds of transfer errors tend to occur in lexical and structural paraphrasing? To find it out, we conducted a preliminary investigation. This section reports a summary of the results. See (Fujita and Inui, 2002) for further details.

We implemented over 28,000 transfer rules for Japanese paraphrases on the KURA paraphrasing engine based on the rules previously reported in (Sato, 1999; Kondo et al., 1999; Kondo et al., 2001; Iida et al., 2001) and existing lexical resources such as thesauri and case frame dictionaries. The implemented rules ranged from such lexical paraphrases as those that replace a word with its synonym to such syntactic/structural paraphrases as those that remove a cleft construction from a sentence, devide a sentence, etc. We then fed KURA with a set of 1,220 sentences randomly sampled from newspaper articles and obtained 630 transferred output sentences.

The following are the tendencies we observed:

- The transfer errors observed in the experiment exhibited a wide range of variety from morphological errors to semantic and discourse-related ones.
- Most types of errors tended to occur regardless of the types of transfer. This suggests that if one creates an error detection module specialized for a particular error type, it works across different types of transfer.
- The most frequent error type involved inappropriate conjugation forms of verbs. It is, however, a matter of morphological generation and can be easily resolved.
- Errors in regard to verb valency and selectional restriction also tended to be frequent and fatal, and thus should have preference as a research topic.
- The next frequent error type was related to the difference of meaning between near synonyms. However, this type of errors could often be detected by a model that could detect errors of verb valency and selectional restriction.

Based on these observations, we concluded that the detection of incorrect verb valences and verb-complement cooccurrence was one of the most serious problems that should have preference as a research topic. We are now conducting experiments on empirical methods for detecting this type of errors (Fujita *et al.*, 2003).

6 Conclusion

This paper reported on the present results of our ongoing research on text simplification for reading assistance targeting congenitally deaf people. We raised four interrelated issues that we needed address to realize this application and presented our previous activities focuing on three of them: readability assessment, paraphrase representation and post-transfer error detection.

Regarding readability assessment, we proposed a novel approach in which we conducted questionnaire surveys to collect readability assessment data and took a corpus-based empirical method to obtain a readability ranking model. The results of the surveys show the potential impact of text simplification on reading assistance. We conducted experiments on the task of comparing the readability of a given paraphrase pair and obtained promising results by SVM-based classifier induction (95% precision with 89% recall). Our approach should be equally applicable to other population segments such as aphasic readers and second-language learners. Our next steps includes the investigation of the drawbacks of the present *bag-of-features* modeling approach. We also need to consider a method to introduce the notion of user classes (e.g. beginner, intermediate and advanced). Textual aspects of readability will also need to be considered, as discussed in (Inui and Nogami, 2001; Siddahrthan, 2003).

Regarding paraphrase representation, we presented our revision-based lexico-structural paraphrasing engine. It provides a fully expressible scheme for representating paraphrases, while preserving the easiness of handcraft paraphrasing rules by providing an extended natural language as a means of pattern editting. We have handcrafted over a thousand transfer rules that implement a broad range of lexical and structural paraphrasing.

The problem of error detection is also critical. When we find a effective solution to it, we will be ready to integrate the technologies into an application system of text simplification and conduct user- and task-oriented evaluations.

Acknowledgments

The research presented in this paper was partly funded by PREST, Japan Science and Technology Corporation. We thank all the teachers at the schools for the deaf who cooperated in our questionnaire survey and Toshihiro Agatsuma (Joetsu University of Education) for his generous and valuable cooperation in the survey. We also thank Yuji Matsumoto and his colleagues (Nara Advanced Institute of Science and Technology) for allowing us to use their NLP tools ChaSen and CaboCha, Taku Kudo (Nara Advanced Institute of Science and Technology) for allowing us to use his SVM tool, and Takaki Makino and his colleagues (Tokyo University) for allowing us to use LiLFeS, with which we implemented KURA. We also thank the anonymous reviewers for their suggestive and encouraging comments.

References

Barzilay, R. and McKeown, K. 2001. Extracting paraphrases from a parallel corpus. In *Proc. of the 39th Annual Meeting and the 10th Conference of the European Chapter of Association for Computational Linguistics (EACL)*, pages 50–57.

Barzilay, R. and Lee, L. 2003. Learning to paraphrases: an unsupervised approach using multiple-sequence alignment. In *Proc. of HLT-NAACL*.

Canning, Y. and Taito, J. 1999. Syntactic simplification of newspaper text for aphasic readers. In *Proc. of the 22nd Annual International ACM SIGIR Conference (SIGIR)*.

Carroll, J., Minnen, G., Canning, Y., Devlin, S. and Tait, J. 1998. Practical simplification of English newspaper text to assist aphasic readers. In *Proc. of AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*.

Dorna, M., Frank, A., Genabith, J. and Emele, M. 1998. Syntactic and semantic transfer with F-structures. In *Proc. of COLING-ACL*, pages 341–347.

Fujita, A. and Inui, K. 2002. Decomposing linguistic knowledge for lexical paraphrasing. In *Information Processing Society of Japan SIG Technical Reports*, NL-149, pages 31–38. (in Japanese)

Fujita, A., Inui, K. and Matsumoto, Y. 2003. Automatic detection of verb valency errors in paraphrasing. In *Information Processing Society of Japan SIG Technical Reports*, NL-156. (in Japanese)

Hermjakob, U., Echihabi, A. and Marcu, D. 2002. Natural language based reformulation resource and Web exploitation for question answering. In *Proc. of the TREC-2002 Conference*.

Iida, R., Tokunaga, Y., Inui, K. and Eto, J. 2001. Exploration of clause-structural and function-expressional paraphrasing using KURA. In *Proc. of the 63th Annual Meeting of Information Processing Society of Japan*, pages 5–6. (in Japanese).

Inui, K. and Nogami, M. 2001. A paraphrase-based exploration of cohesiveness criteria. In *Proc. of the Eighth European Workshop on Natulan Language Generation*, pages 101–110.

Jacquemin, C. 1999. Syntagmatic and paradigmatic representations of term variations. In *Proc. of the 37th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 341–349.

Kondo, K., Sato, S. and Okumura, M. 1999. Paraphrasing of "sahen-noun + suru". *Journal of Information Processing Society of Japan*, 40(11):4064–4074. (in Japanese).

Kondo, K., Sato, S. and Okumura, M. 2001. Paraphrasing by case alternation. *Journal of Information Processing Society of Japan*, 42(3):465–477. (in Japanese).

Kurohashi, S. and Sakai, Y. 1999. Semantic analysis of Japanese noun phrases: a new approach to dictionary-based understanding. In *Proc. of the 37th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 481–488.

Lavoie, B. Kittredge, R. Korelsky, T. Rambow, O. 2000. A framework for MT and multilingual NLG ystems based on uniform lexico-structural processing. In *Proc. of ANLP-NAACL*.

Lin, D. and Pantel, P. 2001. Discovery of inference rules for question-answering. *Natural Language Engineering*, 7(4):343–360.

McCoy ,K. F. and Masterman (Michaud), L. N. 1997. A Tutor for Teaching English as a Second Language for Deaf Users of American Sign Language, In *Proc. of ACL/EACL '97 Workshop on Natural Language Processing for Communication Aids*.

Meyers, A., Yangarber, R. and Grishman, R. 1996. Alignment of shared forests for bilingual corpora. In *Proc. of the 16th International Conference on Computational Linguistics (COLING)*, pages 460–465.

Michaud, L. N. and McCoy, K. F. 2001. Error profiling: toward a model of English acquisition for deaf learners. In *Proc. of the 39th Annual Meeting and the 10th Conference of the European Chapter of Association for Computational Linguistics (EACL)*, pages 386–393.

NIJL, the National Institute for Japanese Language. 1991. *Nihongo Kyôiku-no tame-no Kihon-Goi Chôsa (The basic lexicon for the education of Japanese)*. Shuei Shuppan, Japan. (In Japanese)

Richardson, S., Dolan, W., Menezes, A. and Corston-Oliver, M. 2001. Overcoming the customization bottleneck using example-based MT. In *Proc. of the 39th Annual Meeting and the 10th Conference of the European Chapter of Association for Computational Linguistics (EACL)*, pages 9–16.

Robin, J. and McKeown, K. 1996. Empirically designing and evaluating a new revision-based model for summary generation. *Artificial Intelligence*, 85(1–2):135–179.

Sato, S. 1999. Automatic paraphrase of technical papers' titles. *Journal of Information Processing Society of Japan*, 40(7):2937–2945. (in Japanese).

Shinyama, Y., Sekine, S. Kiyoshi, Sudo. and Grishman, R. 2002. Automatic paraphrase acquisition from news articles. In *Proc. of HLT*, pages 40–46.

Siddahrthan, A. 2003. Preserving discourse structure when simplifying text. In *Proc. of European Workshop on Natural Language Generation*, pages 103–110.

Takahashi, T., Iwakura, T., Iida, R., Fujita, A. and Inui, K. 2001. KURA: a transfer-based lexico-structural paraphrasing engine. In *Proc. of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pages 37–46.

Williams, S., Reiter, E. and Osman, L. 2003. Experiments with discourse-level choices and readability. In *Proc. of European Workshop on Natural Language Generation*, pages 127–134.