

What Prompts Translators to Modify Draft Translations? An Analysis of Basic Modification Patterns for Use in the Automatic Notification of Awkwardly Translated Text

Takeshi Abekawa and Kyo Kageura
Library and Information Science Course
Graduate School of Education,
University of Tokyo
{abekawa, kyo}@p.u-tokyo.ac.jp

Abstract

In human translation, translators first make draft translations and then modify them. This paper analyses these modifications, in order to identify the features that trigger modification. Our goal is to construct a system that notifies (English-to-Japanese) volunteer translators of awkward translations. After manually classifying the basic modification patterns, we analysed the factors that trigger a change in verb voice from passive to active using SVM. An experimental result shows good prospects for the automatic identification of candidates for modification.

1 Introduction

We are currently developing an English-to-Japanese translation aid system aimed at volunteer translators mainly working online (Abekawa and Kageura, 2007). As part of this project, we are developing a module that notifies (inexperienced) translators of awkwardly translated expressions that may need refinement or editing.

In most cases, translators first make draft translations, and then examine and edit them later, often repeatedly. Thus there are normally at least two versions of a given translation, i.e. a draft and the final translation. In commercial translation environments, it is sometimes the case that texts are first translated by inexperienced translators and then edited by experienced translators. However, this does not apply to voluntary translation. In addition, volunteer translators tend to be less experienced than commercial translators, and devote less time to editing. It would therefore be of great help to these translators

if the CAT system automatically pointed out awkward translations for possible modification. In order to realise such a system, it is necessary to first clarify (i) the basic types of modification made by translators to draft translations, and (ii) what triggers these modifications.

In section 2 we introduce the data used in this study. In section 3, we clarify the nature of modification in the translation process. In section 4, we identify the actual modification patterns in the data. In section 5, focusing on “the change from the passive to the active voice” pattern, we analyse and clarify the triggers that may lead to modification. Section 6 is devoted to an experiment in which machine learning methods are used to detect modification candidates. The importance of the various triggers is examined, and the performance of the system is evaluated.

2 The data

The data used in the present study is the Japanese translation of an English book about the problem of peak oil (Leggett, 2005). The book is aimed at a popular audience and is relevant to the sort of texts we have in mind, because the majority of texts volunteer translators translate deal with current affairs, social issues, politics, culture and sports, and/or economic issues for a popular audience¹. The data consists of the English original (henceforth “English”), the draft Japanese translation (“Draft”) and the final Japanese translation (“Final”). The “Draft” was made by two translators (one with two years’ experience and the other with five years’ experience), and

¹Software localisation is another area of translation in which volunteers are heavily involved. We do not include it in our target because it has different characteristics.

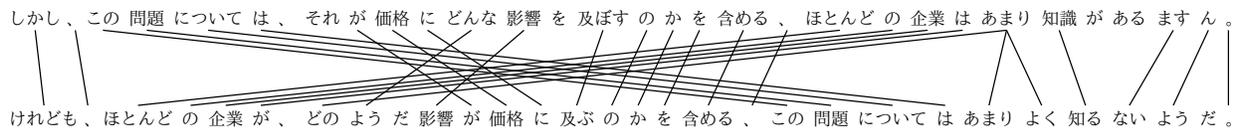


Figure 1: An example of word alignment using GIZA++

the “Final” was made by a translator with 12 years’ experience. Table 1 gives the quantities of the data.

	“English”	“Draft”	“Final”
Number of sentences	4,587	4,629	4,648
Number of words	92,300	127,838	132,989
(Average per sentence)	20.1	27.6	28.6

Table 1: Basic quantities of the data

3 Nature of the modification process

State	Cause
1. Mistranslation	English is complex
2. Text is confusing	English is complex / Translation is too literal
3. Text is unnatural	Translation is too literal / Japanese is underexamined
4. Against modifiers’ taste	Different Japanese ‘model’ is assumed
5. Against editorial policy	Lack of surface editing

Table 2: States in the draft and their causes

As little research has been carried out into the process by which translators modify draft translations, we manually analysed a part of the data in which modifications were made, in consultation with a translator. In the modification process, the translator first recognises (though often not consciously) one of a number of *states* in a draft translation and the underlying cause of the state. S/he then modifies the draft translation if necessary. Table 2 shows the basic classification of *states* and possible *causes*. Although the states are conceptually clear, it is not necessarily the case that translators can judge the state of a given translation consistently, because judging a sentence as being “natural” or “confusing” is not a binary process but a graded one, and the distinction between different states is often not immedi-

ately clear. Many concrete modification patterns found in the data are covered in translation textbooks (Anzai, 1995; Nakamura, 2003). However, although it is obvious in some cases that a section of translated text needs to be modified, in other cases it is less clear, and judgments will vary according to the translator. The task that automatic notification addresses, therefore, is essentially an ambiguous one, even though the actual system output may be binary.

We also identified the distinction between two types of modification: (i) “generative” modification, in which the modified translation is generated on the spot, with reference to the English original; and (ii) “considered” modification, in which alternate expressions (phrases, collocations, etc.) are retrieved from the depository of useful, elegant, or conventional expressions in the translator’s mind. These two types of modification can be activated in the face of one token of modification at once.

4 Modification patterns

The most natural way to classify *modification patterns* is by means of basic linguistic labels such as “change of voice” or “change from nominal modification to adverbial modification” (cf. Anzai, 1995). These modification patterns consist of one or more *primitive operations*. For instance, a “change of voice” may consist of such primitive operations as “changing the case-marker of the subject,” “swapping the position of subject and object,” etc.

As preparation, we extracted modification patterns from the data². In order to do so, we first aligned the “Draft” and the “Final” at the sentence level using DP matching, and then at the morpheme level using GIZA++ (Och and Ney, 2003). Figure 1 illustrates an example of word/morpheme level

²This task is similar to the acquisition of paraphrase knowledge (Barzilay and McKeown, 2001; Shinyama et al., 2002; Quirk et al. 2004; Barzilay and Lee, 2003; Dolan et al., 2004). However, our aim here is to clarify basic modification patterns and not automatic identification.

English:	If it was perceived to be true by the majority of Thinkers, ...				
“Draft”:	人類の JINRUI-NO (thinkers _{genitive})	多数によって TASUU-NIYOTTE (majority _{ablative})	それが SORE-GA (it _{subject})	真実であると SINJITU-DE-ARU-TO (to be true)	認識されれば, NINSIKI-SA-RERE-BA (be perceived)
“Final”:	人類の JINRUI-NO (thinkers _{genitive})	多数が TASUU-GA (majority _{subject})	それを SORE-WO (it _{object})	真実と SINJITU-TO (to be true)	認識すれば, NINSIKI-SURE-BA (perceive)
Primitive operations:		replace(“NIYOTTE”, “GA”)	replace(“GA”, “WO”)	delete(“DE”) delete(“ARU”)	delete(“RARERU”)

Table 3: An example of a primitive modification operation

alignment. Changes in word order occur frequently, as is shown in Figure 1, and the “Final” and the “Draft” are not completely parallel at the word or morpheme level. As a result, GIZA++ sometimes misaligns the units.

From the aligned “Draft” and “Final” data, we identified the primitive operations. We limited these operations to syntactic operations and semantic operations such as the changing of content words, because the latter is hard to generalise with a small amount of data. Primitive operations were extracted by calculating the difference between corresponding *bunsetsu*, which basically consist of a content word and postpositions/suffixes, in the “Draft” and in the “Final”. An example is given in Table 3. Table 4 shows the five most frequent changes in verb inflections and case markers, which are two dominant classes of primitive operation. In addition, we observed deletions and insertions of Sahen verbs.

Modification patterns were identified by observing the degree of co-occurrence among these primitive operations. We used Cabocha³ to identify the syntactic dependencies and used the log-likelihood ratio (LLR) to calculate the degree of co-occurrence of primitive operations that occupy syntactically dependent positions. Table 5 shows the top five pairwise co-occurrence patterns.

inflection	del.	ins.	case marker	del.	ins.
DA	379	291	NI	476	384
TE	269	358	GA	387	502
TA	247	306	NO	366	204
RARERU	224	122	WO	293	421
IRU	197	267	DE	203	193

Table 4: Frequent primitive operations

³<http://chasen.org/~taku/software/cabocha/>

Three main modification patterns were identified: (i) a change from the passive to the active voice (226 cases); (ii) a change from a Sahen verb to a Sahen noun (208 cases); and (iii) a change from nominal modification to clausal structure. These patterns have been discussed in studies of paraphrases (Inui and Fujita, 2004) and in translation textbooks (Anzai, 1995; Nakamura, 2003). We focus on “the change from the passive to the active voice”. It is one of the most important and interesting modification patterns because (i) it is mostly concerned with the main clausal structure in which other modifications are embedded; and (ii) the use of active and passive voices differs greatly between English and Japanese and thus there will be much to reveal.

5 Triggers that lead to modification

Given a draft translation, an experienced translator will be able to recognise any problematic states in it (see Table 2), identify the causes of these states and deal with them. As computers (and inexperienced translators) cannot do the same (cf. Sun et al., 2007), it is necessary to break these causes down into computationally tractable triggers. Keeping in mind the nature of the modification process discussed in section 3, we analysed the actual data, this time with the help of a translator and a linguist.

At the topmost level, two types of triggers were identified: (i) “pushing” triggers that are identified as negative characteristics of the draft translation expressions themselves; and (ii) “pulling” triggers that come from outside (from the depository of expressions in the translator’s mind) and work as concrete “model translations”. The distinction is not entirely clear, because a model is needed in order to identify negative characteristics, and some sort of negative impression is needed for the “model translation” to be called up. The distinction is nevertheless

LLR	f(a,b)	f(a)	f(b)	operation a	operation b	plain expression
146.2	28	35	224	replace(NIYOTTE,GA)	delete(RARERU)	A NIYOTTE B SARERU→A GA B SURU
105.2	34	90	224	replace(GA,WO)	delete(RARERU)	A GA B SARERU→A WO B SURU
91.7	34	115	208	replace(NO,GA)	delete(SAHEN)	A NO B→A GA B SURU
90.9	26	61	208	replace(NO,WO)	delete(SAHEN)	A NO B→A WO B SURU
36.3	15	68	168	replace(NI,WO)	intransitive→transitive	A NI B SURU→A WO C SURU

Table 5: Five of the most frequent co-occurrence patterns between two primitive operations

important, both theoretically and practically. Theoretically, it corresponds to the types of modification observed in section 3. From the practical point of view, the first type is related to the general structural modelling (in its broad sense) of language, while the second is closely related to the status of individual lexicalised expressions. Correspondingly, an NLP system that addresses the first type needs to assume a language model, while a system that addresses the second type needs to call on the relevant external data on the spot. We address the first type of trigger, because we can hypothesise that the modification by change of voice is mainly related to the structural nature of expressions. It should also be noted that, from the machine learning point of view, there are positive and negative features which respectively promote and restrict the modification.

We classified the features that may represent potential triggers into five groups:

(A) Features related to the readability of the English, because the complexity of English sentences (cf. Fry, 1968; Gunning, 1959) can affect the quality of draft translations. Thus the number of words in a sentence, length of words, number of verbs in a sentence, number of commas, etc. can be used as tractable features for automatic treatment.

(B) Features reflecting the correspondence between the English and the draft Japanese translation. Translations that are very literal, either lexically or structurally, are often also awkward. On the other hand, a high degree of word order correspondence can be a positive sign (cf. Anzai, 1995), because it indicates that the information flow in English is maintained and the Japanese translation is well examined.

(C) Features related to the Japanese target verbs. The characteristics of the target verbs should affect the environments in which they occur.

(D) Features related to the “naturalness” of the Japanese. Repetitions or redundancies of elements

or sound patterns may lead to unnatural Japanese sentences.

(E) Features related to the complexity of the Japanese. If a draft translation is too complex, it may be confusing or hard to read. Structural complexity, the length of a sentence, the number of commas, etc. can be used as triggers that reflect the complexity of the Japanese translation.

Table 6 shows the computationally tractable features we defined within this framework. Features with ‘#’ in their name are numeric features and the others are binary features (taking either 0 or 1).

6 Detecting modification candidates

Using these features, we carried out an experiment of automatic identification of modification candidates. As a machine learning method, we used SVM (Vapnik, 1995). The aim of the experiment was twofold: (i) to observe the feasibility of automatic notification of modification candidates, and (ii) to examine the factors that trigger modifications in more detail.

6.1 Experimental setup

In the application of SVM, we reduced the number of binary features by using those that have higher correlations with positive and negative examples, using mutual information (MI). Table 7 shows features that have high correlations with positive and negative features (eight for each).

SVM settings: The liner kernel was used. For a numeric feature X , the value x is normalized by z-score, $norm(x) = \frac{x-avg(X)}{\sqrt{var(X)}}$, where $avg(x)$ is the empirical mean of X and $var(X)$ is the variance of X .

Data: The numbers of positive and negative cases in the data are 226 and 894, respectively (1120 in total). In order to balance the positive and negative examples, we used an equal number of examples for training.

(A)			
$EN_{\#word}$:	the number of words in the English sentence	S_{first_agent} :	first case element in the sentence has an AGENT attribute
$EN_{\#pause}$:	the number of delimiters in the English sentence	$S_{before_passive}$:	Is there a passive verb before the target verb in the sentence?
$EN_{\#verb}$:	the number of verbs in the English sentence	$S_{after_passive}$:	Is there a passive verb after the target verb in the sentence?
$EN_{\#VNN}$:	the number of VNN verbs in the English sentence	(D)	
$EN_{\#word_len}$:	the average number of characters in a word	$N_{modifying_voice}$:	the voice of the verb that modifies the target verb
(B)		$N_{modifying_voice}$:	the voice of the verb that is modified by the target verb
E_{POS} :	POS of the English word corresponding to the target Japanese verb	$N_{grandparent_voice}$:	the voice of the grandparent verb of the target verb
E_{POS_before} :	POS of a word before the English word corresponding to the target Japanese verb	$N_{grandchild_voice}$:	the voice of the grandchild verb of the target verb
E_{POS_after} :	POS of a word after the English word corresponding to the target Japanese verb	$N_{case_adjacency}$:	bigram consists of a particle of the target verb and a particle of the adjacency bunsetsu chunk
$E_{POS_before:POS}$:	a bigram of E_{POS_before} and E_{POS}	(E)	
$E_{POS:POS_after}$:	a bigram of E_{POS} and E_{POS_after}	$J_{\#morpheme}$:	the number of morphemes in the target Japanese sentence
$EJ_{\#translation}$:	translation probability between the source and target language sentences	$J_{\#pause}$:	the number of pause marks in the target Japanese sentence
(C)		$J_{\#verb}$:	the number of verbs in the target Japanese sentence
F_{suffix} :	a suffix following the target verb	$J_{\#passive}$:	the number of verbs with passive voice in the target Japanese sentence
$F_{particle}$:	a particle following the target verb	$J_{\#depth}$:	depth of the modifier which modifies the target verb
F_{pause_mark} :	a pause mark following the target verb		
$D_{modifying_case}$:	case marker of the element that modifies the target verb		
$D_{modifying_agent}$:	case marker of the element that modifies the target verb, if its case element has an AGENT attribute		
$D_{functional}$:	functional noun which is modified by the target verb		
$D_{modified_case}$:	case marker of the element that is modified by the target verb		

Table 6: Features

Methods of evaluation: We used (i) 10-fold cross validation to check the power of classifiers for unknown data and (ii) a partially closed test in which the 226 positive and negative examples were used for training and 1120 data were evaluated, in order to observe the realistic prospects for actual use.

6.2 Result of experiment and feature analysis

Table 8 shows the results. Though they are reasonable, the overall accuracy, especially for the partially closed test, shows that the method is in need of improvement.

In order to evaluate the effectiveness of the feature sets, we carried out experiments only using and without using each feature set. Table 9 shows that how efficient is each feature set defined in Table 6. The left-hand column in Table 9 shows the result with all feature sets except focal feature set, and the right-hand column shows the result when only the

focal feature set was used.

The experiment showed that the feature set that contributed most was C (features related to the Japanese target verbs). We also carried out an experiment to check which features are effective among this set, in the same manner as the experiments for checking the effectiveness of the feature sets. The result showed that the feature $D_{modifying_case}$ is the feature that contributed the most by far. In Japanese, case markers are strongly correlated with the voice of verbs, and the coverage of this feature for tokens related to voice is high because it is common for a verb to be modified by the case element with the case marker.

It became clear that the numeric features A and E contribute little to the overall accuracy. Table 10 shows the correlation coefficient between the numeric features and correct answers. The table shows that there is no noticeable relation between the nu-

	accuracy	(+)precision	(+)recall	(-)precision	(-)recall
Cross validation	0.646 (291/452)	0.656 (138/214)	0.614 (138/226)	0.643 (153/238)	0.677 (153/226)
Partially closed	0.521 (583/1120)	0.277 (193/697)	0.854 (193/226)	0.922 (390/423)	0.436 (390/894)

Table 8: The accuracy of classification

feature set	without this feature set			using only this feature set		
	accuracy	(+)precision	(+)recall	accuracy	(+)precision	(+)recall
(A)	0.638	0.638 (144/226)	0.639 (144/226)	0.521	0.541 (62/115)	0.277 (62/226)
(B)	0.634	0.649 (132/203)	0.584 (132/226)	0.563	0.549 (159/290)	0.705 (159/226)
(C)	0.579	0.576 (136/237)	0.604 (136/226)	0.610	0.620 (128/207)	0.570 (128/226)
(D)	0.645	0.654 (138/212)	0.615 (138/226)	0.523	0.679 (19/29)	0.087 (19/226)
(E)	0.629	0.666 (117/175)	0.518 (117/226)	0.492	0.491 (101/205)	0.447 (101/226)

Table 9: The evaluation result for each feature set

feature	MI	f(+)	f(-)	feature set (A)	
$D_{modifying_agent}=NIYOTTE$	0.843	15	17	$EN_{\#word}$	0.038
$E_{POS:POS_after}=VVN:NN$	0.656	14	22	$EN_{\#pause}$	-0.069
$E_{POS_before}=IN$	0.536	10	19	$EN_{\#verb}$	-0.003
$E_{POS_before}=JJ$	0.530	12	23	$EN_{\#VVN}$	-0.061
$D_{modified_case}=GA$	0.428	13	29	$EN_{\#word_len}$	0.033
$N_{grandparent_voice}=passive$	0.408	17	39	feature set (E)	
$N_{grandchild_voice}=passive$	0.368	14	34	$J_{\#morpheme}$	0.083
$E_{POS}=VVZ$	0.368	14	34	$J_{\#pause}$	0.011
$F_{suffix}=NARU$	0.225	0	23	$J_{\#verb}$	0.056
$N_{case_adjacency}=GA:TO$	0.225	0	12	$J_{\#passive}$	0.035
$F_{suffix}=SHIMAU$	0.225	0	16	$J_{\#depth}$	0.098
$E_{POS}=RB$	0.225	0	10		
$E_{POS:POS_after}=VVG:DT$	0.225	0	10		
$E_{POS:POS_after}=VVN:TO$	0.179	2	42		
$E_{POS:POS_after}=VVN:SENT$	0.159	3	44		
$D_{modifying_agent}=NI$	0.154	4	54		

Table 7: Features which have high correlation with positive and negative examples

meric features and the correct results. We introduced most numeric features based on the study of readability. In readability studies, however, these features are defined in terms of the overall document, and not in terms of individual sentences or of verb phrases. It would be preferable to develop numerical features that can properly reflect the nature of individual sentences or smaller constructions.

Table 9 shows that the result when only using the feature set D has a very low recall, but the highest

Table 10: The correlation coefficient between each feature and correct answer

precision of all the feature sets. This mean that there are not many occasions on which the feature set D can be applied, but when it is applied, the result is reliable. The feature set D thus is efficient as a trigger once it is applied, and the different treatment of the tokens that contain this feature set may contribute to the performance improvement.

6.3 Diagnosis

The critical cases from the point of view of improving the performance are the false positives and false negatives. We thus manually analysed the false positives and false negatives obtained in the partially closed experiment (in the actual application environment, as much training data as available should be used; we thus used the results of the partially

closed experiment here). For the false positive, we extracted 100 sample sentences from 504 sentences. For the false negative we used all 33 sentences. We asked two translators to judge whether (i) it would be better to modify the draft translations or (ii) it would not be necessary to modify the draft translations.

6.3.1 False positives

From the 100 sample sentences, we excluded 23 cases, 18 of which were judged as in need of modification by one of the translators and 5 of which were judged as in need of modification by both of the translators. We manually analysed the remaining 77 cases. Rather than the problems with the features that we used, we identified the potential factors that would contribute to the restriction of modification. Three types of restricting factor were recognised:

1. The nature of individual verbs allows or requires the passive voice. Within the data, three subtypes were identified, i.e. (i) the use of the passive is natural irrespective of context, as in “消費され (consumed)” (48 cases); (ii) the use of the passive is natural within certain fixed syntactic patterns, as in “X と呼ばれる Y (Y called X)” (10 cases); and (iii) the passive is used as part of a common collocation, as in “不安に襲われた (attacked by anxiety)” (2 cases);
2. The use of the active voice is blocked by selectional restrictions, as in “作られた堆積物 (a sediment made by ...)” (1 case); and
3. The structure of the sentence requires the passive, as in “最大の企業はすべて車を製造する企業であり、その中で石油の大半が消費されていた (The biggest companies were all companies making cars, in which most of the oil was consumed)” (16 cases).

Together they cover 73 cases (in 4 out of 77 cases we could not identify the factor, and in 4 of the 73 cases two of the above factors were identified). It is anticipated that the first type (60 cases; about 85%) could be dealt with by introducing “pulling” triggers, i.e. using large corpora to identify the characteristics of the use of voice for individual verbs, in order to enable the system to judge the desirability of given expressions vis-à-vis the conventional alternatives. To deal with the second type requires a detailed semantic description of nouns, which is

difficult to achieve, though in some cases it could be approximated by collocational tendencies. In regards to the third type of false positive, we expected that the type of features used in the experiment would have been sufficient to eliminate them, but this was not the case. In fact, many of the features require discourse level information, such as the choice of subject within the flow of discourse, in order to function properly, which we did not take into account. Although high-performance discourse processing is still in an embryonic stage, in the setting of the present study the correspondence between key information in English and that in Japanese could be used to deal with this type of false positive.

6.3.2 False negatives

Here, it is necessary to find factors that would promote modification. Among the 33 false negatives, 4 were judged as not in need of modification by both the translators. We thus examined the remaining 29 cases. In 13 cases, the verb was replaced by another verb. Including these cases, we identified four basic factors that are related to triggering modification:

1. The nature of the individual verbs strongly requires the active voice, either independently or within the particular context, as in “から尋ねられました (was asked by)” (9 cases);
2. The structure of the sentence is rendered rather awkward by the use of passives, as in “に発表されたアナリストたちによるレポートである (a report published in by analysts)” (4 cases);
3. A given lexical collocation is unnatural or awkward, as in “すべての投資がふるいにかけることを共同で要求し (that all investments be screened is collectively insisted)” (2 cases); and
4. A lexicalised collocation in the draft was subtly awkward and there is a better collocation or expression that fits the situation (14 cases).

Together they cover 26 cases. We could not identify features in 3 cases. As in false positives, the first, second and fourth types (22 cases or about 85% are fully covered by these three types) could be dealt with by introducing “pulling” triggers, using large external corpora.

For the overall data, we would expect that around 85% of 388 (77% of 504 cases) false positives (330

cases) could be dealt with by introducing “pulling” triggers. If these false positives could be removed completely, the precision would become well over 0.5 (193/(697-330)) and the ratio of notified cases would become about one third ((697-330)/1120) of the total relevant cases. Though it is unreasonable to assume this ideal case, this indicates that the features we defined and introduced in this study — though limited to those related to “pushing” triggers — were effective, and that what we have achieved by using these features is very promising in terms of realising a system that notifies users of awkward translations.

7 Conclusions

In this paper, we examined the factors that trigger modifications when translators are revising draft translations, and identified computationally tractable features relevant to the modification. We carried out an experiment for automatic detection of modification candidates. The result was highly promising, though it revealed several issues that need to be addressed further.

Following the results reported in this paper, we are currently working on.

- (i) extending the experiment by introducing outside data to carry out open experiments (we have obtained draft and final translations of three more books);
- (ii) introducing the degree of necessity for modifications by asking translators to judge the data; and
- (iii) further examining the features used in the experiment for the improvement of performance.

In addition, we are experimenting with a method for making use of large-scale external corpora in order to deal with “pulling”-type triggers, with additional features taken from large external corpora.

Acknowledgement

This research is partly supported by grant-in-aid (A) 17200018 “Construction of online multilingual reference tools for aiding translators” by the Japan Society for the Promotion of Sciences (JSPS), and also by grant-in-aid from The HAKUHO FOUNDATION, Tokyo.

References

- Abekawa, T. and Kageura, K. 2007. A translation aid system with a stratified lookup interface. In *Proc. of ACL 2007 Demos and Poster Sessions*, p. 5–8.
- Anzai, T. 1995. *Eibun Hon'yaku Jutu* (in Japanese). Tokyo: Chikuma.
- Barzilay, R. and McKeown, K. R. 2001. Extracting paraphrases from a parallel corpus. In *Proc. of ACL 2001*, p. 50-57.
- Barzilay, R. and Lee, L. 2003. Learning to paraphrase: An unsupervised approach using multiple-sequence alignment. In *Proc. of HLT-NAACL 2003*, p. 16-23.
- Dolan, B. et. al. 2004. Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources alignment. In *Proc. of COLING 2004*, p. 350-356.
- Fry, E. 1968. A readability formula that saves time. *Journal of Reading*, 11, p. 513-516, 575-578.
- Gunning, R. 1959. *The Technique of Clear Writing*. New York: McGraw-Hill.
- Haruno, M. and Yamazaki, T. 1996. High-performance bilingual text alignment using statistical and dictionary information. In *Proc. of ACL 1996*, p. 131-138.
- Inui, K. and Fujita, A. 2004. A survey on paraphrase generation and recognition. *Journal of Natural Language Processing*, 11(5), p. 131-138.
- Leggett, J. 2005. *Half Gone*. London: Portobello. [Masuoka, K. et. al. trans. 2006. *Peak Oil Panic*. Tokyo: Sakuhinsha.]
- Nakamura, Y. 2003. *Eiwa Hon'yaku no Genri Gihou* (in Japanese). Tokyo: Nichigai Associates.
- Och, F. J. and Ney, H. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1), p. 19-51.
- Quirk, C., Brocktt, C. and Dolan, W. B. 2004 Monolingual machine translation for paraphrase generation. In *Proc. of EMNLP 2004*, p. 142-149.
- Schmid, H. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proc. of NeMLAP*, p. 44-49.
- Shinyama, Y. et. al. 2002. Automatic paraphrase acquisition from news articles. In *Proc. of HLT 2002*, p. 40-46.
- Sun, et. al. 2007. Detecting erroneous sentences using automatically mined sequential patterns. In *Proc. of ACL 2007*, p. 81-88.
- Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. New York: Springer.