Zero Pronoun Resolution can Improve the Quality of J-E Translation

Hirotoshi Taira, Katsuhito Sudoh, Masaaki Nagata
NTT Communication Science Laboratories
2-4, Hikaridai, Seika-cho, Keihanna Science City
Kyoto 619-0237, Japan
{taira.hirotoshi,sudoh.katsuhito,nagata.masaaki}@lab.ntt.co.jp

Abstract

In Japanese, particularly, spoken Japanese, subjective, objective and possessive cases are very often omitted. Such Japanese sentences are often translated by Japanese-English statistical machine translation to the English sentence whose subjective, objective and possessive cases are omitted, and it causes to decrease the quality of translation. We performed experiments of J-E phrase based translation using Japanese sentence, whose omitted pronouns are complemented by human. We introduced 'antecedent F-measure' as a score for measuring quality of the translated English. As a result, we found that it improves the scores of antecedent F-measure while the BLEU scores were almost unchanged. Every effectiveness of the zero pronoun resolution differs depending on the type and case of each zero pronoun.

1 Introduction

Today, statistical translation systems have been able to translate between languages at high accuracy using a lot of corpora . However, the quality of translation of Japanese to English is not high comparing with the other language pairs that have the similar syntactic structure such as the French-English pair. Particularly, the quality of translation from spoken Japanese to English is in low. There are many reasons for the low quality. One is the different syntactic structures, that is, Japanese sentence structure is SOV while English one is SVO. This problem has been partly solved by head finalization techniques (Isozaki et al., 2010). Another big problem is that subject, object and possessive cases are often eliminated in Japanese, particularly, spoken Japanese (Nariyama, 2003). In the case of Japanese to English translation, the source language has lesser information in surface than the target language, and the quality of the translation tends to be low. We show the example of the omissions in Fig 1. In this example, the Japanese subject *watashi wa* ('I') and the object *anata ni* ('to you') are eliminated in the sentence. These omissions are not problems for human speakers and hearers because people easily recognize who is the questioner or responder (that is, 'I' and 'you') from the context. However, generally speaking, the recognition is difficult for statistical translation systems.

Some European languages allow the elimination of subject. We show an example in Spanish in Fig 2. In this case, the subject is eliminated, and it leaves traces including the case and the sex, on the related verb. The Spanish word, *tengo* is the first person singular form of the verb, *tener* (it means 'have'). So it is easier to resolve elimination comparing with Japanese one for SMT.

Otherwise, Japanese verbs usually have no inflectional form depending on the case and sex. So, we need take another way for elimination resolution. For example, if the eliminated Japanese subject is always 'I' when the sentence is declarative, and the subject is always 'you' when the sentence is a question sentence, phrase based translation systems are probably able to translate subject-eliminated Japanese sentences to correct English sentences. However, the hypothesis is not always

Proceedings of SSST-6, Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation, pages 111–118,
Jeju, Republic of Korea, 12 July 2012.

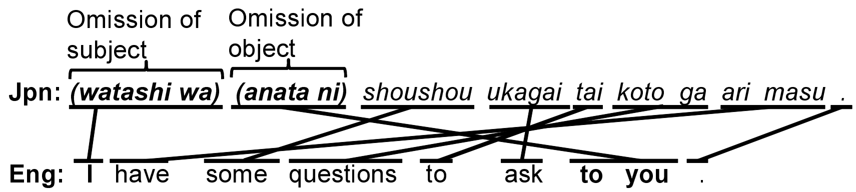

Figure 1: Example of Japanese Ellipsis (Zero Pronoun)

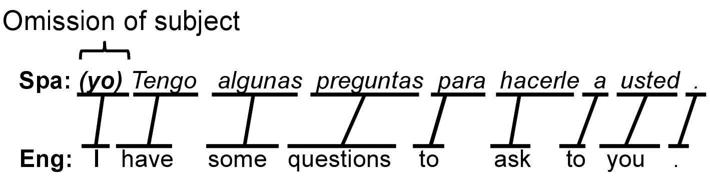

Figure 2: Spanish Ellipsis

true.

In this paper, we show that the quality of spoken Japanese to English translation can improve using a phrase-based translation system if we can use an ideal elimination resolution system. However, we also show that a simple elimination resolution system is not effective to the improvement and it is necessary to recognize correctly the modality of the sentence.

2 Previous Work

There are a few researches for adaptation of ellipsis resolution to statistical translation systems while there are a lot of researches for one to rule-based translation systems in Japanese (Yoshimoto, 1988; Dohsaka, 1990; Nakaiwa and Yamada, 1997; Yamamoto et al., 1997).

As a research of SMT using elimination resolution, we have (Furuichi et al., 2011). However, the target of the research is illustrative sentences in English to Japanese dictionary. Our research aims spoken language translation and it is different from the paper.

3 Setup of the Data of Subjects and Objects Ellipsis in Spoken Japanese

3.1 Ellipsis Resolved Data by Human

In this section, we describe the data used in our experiments. We used BTEC (Basic Travel Expression Corpus) corpus (Kikui et al., 2003) distributed in IWSLT07 (Fordyce, 2007). The corpus consists of tourism-related sentences similar to those that are usually found in phrasebooks for tourists going abroad. The characteristics of the dataset are shown in Table 1. We used 'train' for training, 'devset1-3' for tuning, and 'test' for evaluation. We did not use the 'devset4' and 'devset5' sets because of the different number of English references.

We annotated zero pronouns and the antecedents to the sentences by hand. Here, zero pronoun is defined as an obligatory case noun phrase that is not expressed in the utterance but can be understood through other utterances in the discourse, context, or out-of-context knowledge (Yoshimoto, 1988). We annotated the zero pronouns based on pronouns in the translated English sentences. The BTEC corpus has multi-references in English. We first chose the most syntactically and lexically similar translation in the references and annotated zero pronouns in it. Our target pronouns are *I, my, me, mine, myself, we, our, us, ours, ourselves, you, your, yourself, yourselves, he, his, him, himself, she, her, herself, it, its, itself, they, their, them, theirs and themselves* in English. We show the distribution of the annotation types in the test set in Table 2.

3.2 Baseline System

We also examined a simple baseline zero pronoun resolution system for the same data. We defined

Table 1: Data distribution

	train	devset1-3	devset4	devset5	test
# of References	1	16	7	7	16
# of Source Segments	39,953	1,512	489	500	489

Japanese predicate as verb, adjective, and copula (*da* form) in the experiments. If the inputted Japanese sentence contains predicates and it does not contain 'wa' (a binding particle and a topic marker), 'mo' (a binding particle, which means 'also' and can often replace 'wa' and 'ga'), and 'ga' (a case particle and subjective marker), the system regards the sentence as a candidate sentence to solve the zero pronouns. Then, if the candidate sentence is declarative, the system inserts '*watashi wa* (I)' when the predicate is a verb, and '*sore wa* (it)' when the predicate is a adjective or a copula. In the same way, if the candidate sentence is a question, the system inserts '*anata wa* (you)' when the predicate is a verb, and '*sore wa* (it)' when the predicate is a adjective or a copula. These inserted position is the beginning of the sentence. In the case that the sentence is imperative, the system does not solve the zero pronouns (Fig. 3).

4 Experiments

4.1 Experimental Setting

Fig. 4 shows the outline of the procedure of our experiment. We used Moses (Koehn et al., 2007) for the training of the translation and language models, tuning with MERT (Och, 2003) and the decoding. First, we prepared the data for learning which consists of parallel English and Japanese sentences. We used MeCab [1] as Japanese tokenizer and the tokenizer in Moses Tool kit as English tokenizer. We used default settings for the parameters of Moses. Next, Moses learns language model and translation model from the Japanese and English sentence pairs. Then, the learned model was tuned by completed sentences with MERT. and Moses decoded the completed Japanese sentences to English sentences.

4.2 Evaluation Method

We used BLEU (Papineni et al., 2002) and antecedent Precision, Recall and F-measure for the evaluation of the performances, comparing the system outputs with the English references of test data. Using only BLEU score is not adequate for evaluation of pronoun translation (Hardmeier et al., 2010).

We were inspired empty node recovery evaluation by (Johnson, 2002) and defined antecedent Precision (P), Recall (R) and F-measure (F) as follows,

$$P = \frac{|G \cap S|}{|S|}$$

$$R = \frac{|G \cap S|}{|G|}$$

$$F = \frac{2PR}{P + R}$$

Here, S is the set of each pronoun in English translated by decoder, G is the set of the gold standard zero pronoun.

We evaluated the effect of performance of every case among completed sentences by human, ones by the baseline system, and the original sentences.

4.3 Experimental Result

We show the BLEU scores in Table 3. and the antecedent precision, recall and F-measure in Table 4. The BLEU scores for experiments using our baseline system and human annotation, are slightly better than for one without ellipsis resolution, 45.4% and 45.6%, respectively. However, the scores of antecedent F-measure have major difference between 'original' and 'human'. Particularly, the recall is improved. Each 1st, 2nd and 3rd person score is better than original one.

5 Discussion and Conclusion

We performed experiments of J-E phrase based translation using Japanese sentences, whose omitted pronouns are complemented by human and a baseline system. Using 'antecedent F-measure' as a score for measuring the quality of the translated English, it improves the score of antecedent F-measure. Every effectiveness of the zero pronoun resolution

[1] http://mecab.sourceforge.net/

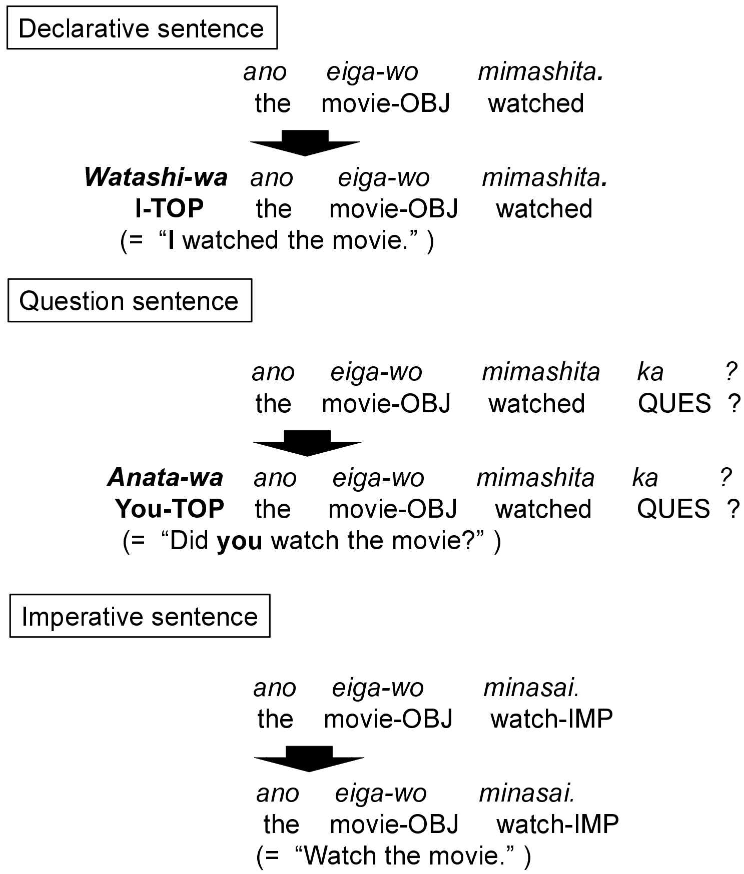

Figure 3: Our baseline system of zero pronoun resolution

differed, depending on the type and case of each zero pronoun. The F-measures for the first person pronoun were smaller than expected ones, Rather, the scores for and possessive pronouns second person were greater (Table. 3).

We show a better, a worse, and an unchanged cases of translation using the baseline system of the elimination resolution in Fig. 5. The left-hand is the result of the alignment between the original Japanese sentence and the decoded English sentence. The right-hand is the result of one using the Japanese the baseline system solved zero pronouns. In the 'better' case, the alignment of *todokete* (send) is better than one of the original sentence, and 'Can you' is compensated by the solved zero pronoun *anata-wa* (you-TOP). Otherwise, in the 'worse' case, our baseline system could not recognize that the sentence is imperative, and inserted *watashi-wa* (I-TOP) incorrectly into the sentence. It indicates that we need a highly accurate recognition of the modalities of sentences for more correct completion of the antecedent of zero pronouns. In the 'unchanged' case, the translation results are the same. However, the alignment of the right-hand is more correct than one of the left-hand.

References

Kohji Dohsaka. 1990. Identifying the referents of zero-pronouns in japanese based on pragmatic constraint interpretation. In *Proceedings of ECAI*, pages 240–245.

C.S. Fordyce. 2007. Overview of the iwslt 2007 evaluation campaign. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 1–12.

M. Furuichi, J. Murakami, M. Tokuhisa, and M. Murata. 2011. The effect of complement subject in japanese to english statistical machine translation (in Japanese). In *Proceedings of the 17th Annual Meeting of The*

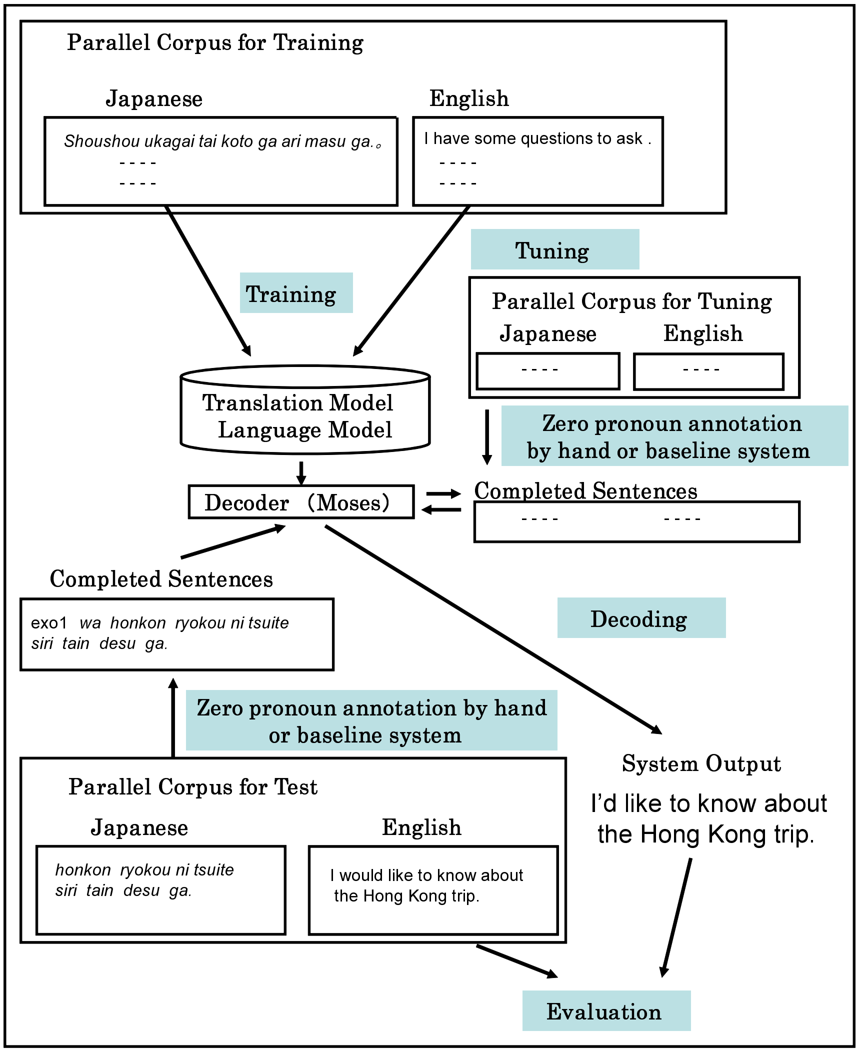

Figure 4: Outline of the experiment

Association for Natural Language Processing (NLP-2012).

C. Hardmeier, M. Federico, and F.B. Kessler. 2010. Modelling pronominal anaphora in statistical machine translation. In *Proceedings of the seventh International Workshop on Spoken Language Translation (IWSLT)*, pages 283–289.

H. Isozaki, K. Sudoh, H. Tsukada, and K. Duh. 2010. Head finalization: A simple reordering rule for sov languages. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and Metrics-MATR*, pages 244–251. Association for Computational Linguistics.

Mark Johnson. 2002. A simple pattern-matching algorithm for recovering empty nodes and their antecedents. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 136–143, Philadelphia, Pennsylvania, USA, July. Association for Computational Linguistics.

G. Kikui, E. Sumita, T. Takezawa, and S. Yamamoto. 2003. Creating corpora for speech-to-speech translation. In *Proceedings of EUROSPEECH*, pages 381–384.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: open source toolkit for statistical machine translation. In *Proc. of the 45th Annual Conference of the Association for*

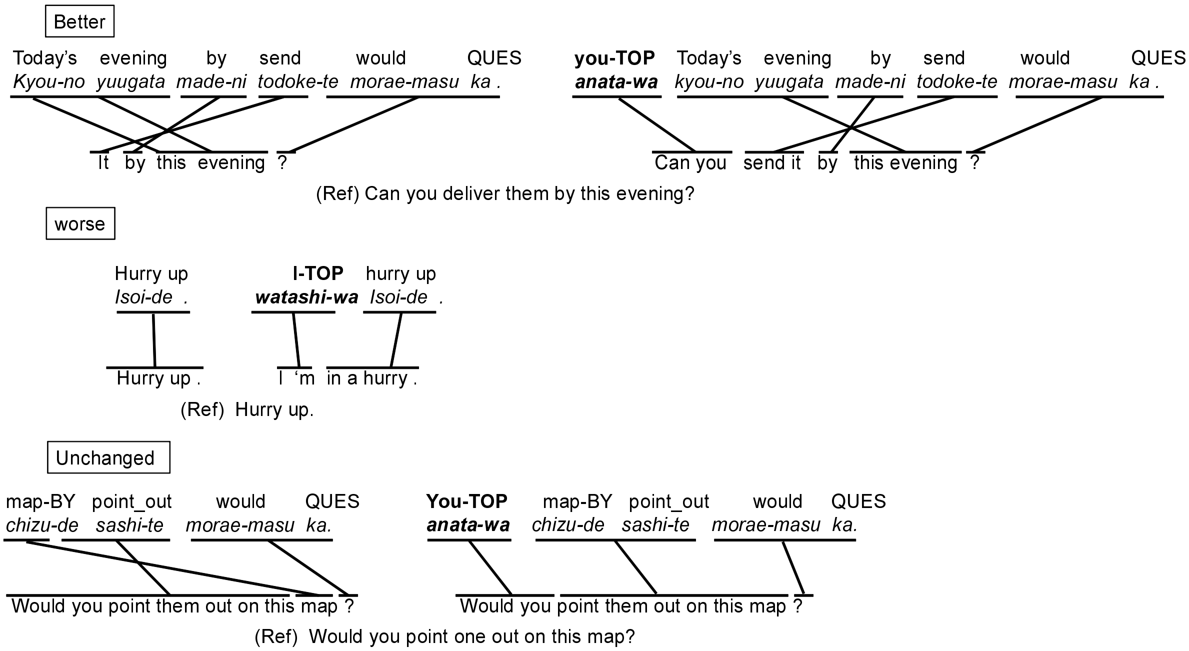

Figure 5: Effectiveness of zero pronoun resolution for decoding

Computational Linguistics (ACL-07), Demonstration Session, pages 177–180.

H. Nakaiwa and S. Yamada. 1997. Automatic identification of zero pronouns and their antecedents within aligned sentence pairs. In *Proc. of the 3rd Annual Meeting of the Association for Natural Language Processing*.

S. Nariyama. 2003. *Ellipsis and reference tracking in Japanese*, volume 66. John Benjamins Publishing Company.

Franz Josef Och. 2003. Minimum error rate training for statistical machine translation. In *Proc. of the ACL*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proc. of the 40th Annual Conference of the Association for Computational Linguistics (ACL-02)*.

K. Yamamoto, E. Sumita, O. Furuse, and H. Iida. 1997. Ellipsis resolution in dialogues via decision-tree learning. In *Proc. of NLPRS*, volume 97. Citeseer.

K. Yoshimoto. 1988. Identifying zero pronouns in japanese dialogue. In *Proceedings of the 12th conference on Computational linguistics-Volume 2*, pages 779–784. Association for Computational Linguistics.

Table 2: The Type Distributions of Zero Pronouns in Test Set

Type	Pronoun	#
First personal pronoun	i	121
	my	39
	me	32
	mine	1
	myself	0
	we	7
	our	2
	us	2
	ours	0
	ourselves	0
	total	204
Second personal pronoun	you	95
	your	23
	yours	0
	yourself	0
	yourselves	0
	total	118
Third personal pronoun	he	1
	his	0
	him	0
	himself	0
	she	0
	her	2
	hers	0
	herself	0
	it	51
	its	0
	itself	0
	they	2
	their	0
	them	5
	theirs	0
	themselves	0
	total	61
all	total	383

Table 3: BLEU score

	BLEU	F(Avg.)	P	R	F (1st person)	F (2nd person)	F (3rd person)
original	45.1	59.7	63.8	56.1	61.6	59.9	52.3
baseline	45.4	58.5	64.1	53.7	61.2	59.2	47.7
human	**45.6**	**71.8**	**67.5**	**76.7**	**70.6**	**77.6**	**63.7**

Table 4: Antecedent precision, recall and F-measure for every pronoun

		i (ref:121)			my (ref:39)			me (ref:32)		
	BLEU	P	R	F	P	R	F	P	R	F
original	45.1	**56.8**	51.2	53.9	55.5	51.2	53.3	58.0	56.2	57.1
baseline	45.4	51.8	46.2	48.9	**67.8**	48.7	56.7	**66.6**	50.0	57.1
human	**45.6**	50.9	**68.6**	**58.4**	65.2	**76.9**	**70.5**	61.2	**59.3**	**60.3**

	we (ref:7)			our (ref:2)			us (ref:2)		
	P	R	F	P	R	F	P	R	F
original	20.0	14.2	16.6	100.0	50.0	66.6	0.00	0.00	0.00
baseline	25.0	14.2	18.1	100.0	50.0	66.6	0.00	0.00	0.00
human	**40.0**	**28.5**	**33.3**	100.0	50.0	66.6	0.00	0.00	0.00

	you (ref:95)			your (ref:23)		
	P	R	F	P	R	F
original	55.3	54.7	55.0	**80.0**	52.1	63.1
baseline	57.1	54.7	55.9	58.8	43.4	50.0
human	**68.4**	**80.0**	**73.7**	73.0	**82.6**	**77.5**

	it (ref:51)			its (ref:0)		
	P	R	F	P	R	F
original	56.1	45.1	50.0	0.00	0.00	0.00
baseline	51.2	41.1	45.6	0.00	0.00	0.00
human	**58.3**	**54.9**	**56.5**	0.00	0.00	0.00

	they (ref:2)			their (ref:0)			them (ref:5)		
	P	R	F	P	R	F	P	R	F
original	**100.0**	50.0	**66.6**	0.00	0.00	0.00	0.00	0.00	0.00
baseline	**100.0**	50.0	**66.6**	0.00	0.00	0.00	0.00	0.00	0.00
human	58.3	**54.9**	56.5	0.00	0.00	0.00	0.00	0.00	0.00