Noun phrase reference
in Japanese-to-English machine translation

Francis BOND, Kentaro OGURA and Tsukasa KAWAOKA*

NTT Communication Science Laboratories

1-2356 Take. Yokosuka-shi, Kanagawa-ken, JAPAN 238-03

bond@nttkb.ntt.jp

Abstract

This paper shows the necessity of distinguishing different referential uses of noun phrases in machine translation. We argue that differentiating between the generic, referential and ascriptive uses of noun phrases is the minimum necessary to generate articles and number correctly when translating from Japanese to English. Heuristics for determining these differences are proposed for a Japanese-to-English machine translation system. Finally the results of using the proposed heuristics are shown to have raised the percentage of noun phrases generated with correct use of articles and number in the Japanese-to-English machine translation system **ALT-J/E** from 65% to 77%.

1   Introduction

Determining the referential property of noun phrases is essential not only to understanding a text, but also to decide how to generate it in English. This paper proposes a heuristic algorithm to determine the referential properties of noun phrases in a Japanese text. The original motivation of the research was to improve the quality of English output by NTT Communication Science Laboratories' Japanese to English machine translation system **ALT-J/E** (Ikehara et al., 1991; Ogura et al., 1993). We expect, however, that the results will also be useful for text extraction and general text understanding.

In this paper we use the term **noun phrase reference** to describe the relation between a noun phrase and what it stands for when it is used. We distinguish between three uses of noun phrases, two referential and one non-referential. A noun phrase can be used to refer in two different ways: GENERIC where a noun phrase is used to refer to a whole class, and REFERENTIAL where a noun phrase refers to a particular entity or entities. A third use is ASCRIPTIVE where a noun

*Now at **Dōshisha** University, Kyoto, JAPAN: <kawaoka@wise.doshisha.ac.jp>.

phrase is used not to refer to anything but rather, normally with a copula verb, to ascribe a property to some referent. Although ASCRIPTIVE noun phrases are non-referring, we will refer to all three uses under the general term of noun phrase reference. This three-way distinction of noun phrase reference was introduced in Bond et al. (1994) and used as a, base to determine the countability and number of noun phrases in Japanese-to-English machine translation. In this paper we define exactly what is meant by the three kinds of reference and show how the distinction is essential in the generation of articles.

This paper is structured as follows. First, we define the three kinds of referentiality which we distinguish and justify the definitions on theoretical and practical grounds, comparing them with those suggested by other researchers. We then describe in detail a heuristic method for determining noun phrase reference in Japanese sentences. Next, we show how the distinction is used in a Japanese to English machine translation system to generate articles and number. Finally, we look at experimental results gained by implementing the proposed methods and compare them to those achieved by an earlier version of the same system, and by other systems.

2   Definition of noun phrase reference

Noun phrase reference is of fundamental importance in any discussion of meaning (Lyons 1977). In English, it is also important in determining how articles should be used. In this section we give a more detailed definition of the three kinds of noun phrase reference under discussion and compare them with the definitions used in other machine translation systems.

Generic: Noun phrases with generic reference denote an entire class: e.g. *mammoth* in *Mammoths are extinct.* In English generic noun phrases can normally be expressed in three ways, as discussed in Section 4.1.

Referential: Referential noun phrases are those that refer to some entity or entities in the discourse world: e.g. *mammoth* in *There is a mammoth in my garden!* Referential noun phrases are plural if there is more than one discrete referent, and are marked for definiteness.

Ascriptive: Ascriptive noun phrases are used with a copula verb, or in an appositive expression, to ascribe a property to their subject: e.g. *a mammoth* in *That animal is a mammoth.* Because ascriptive noun phrases are non-referring they cannot be the antecedent of other noun phrases.

Zelinsky-Wibbelt (1992) distinguishes between GENERIC and IDENTIFYING, which appear to be equivalent to our GENERIC and REFERENTIAL. Zelinsky-Wibbelt's examples do contain ascriptive noun phrases, for example *a human being* in *'A spectator is a human being',* instead they appear to be treated as adjective phrases in the rules (for example in their rule 14 (p. 797 op cit) where the

complement of the copulative predicate with a generic subject is an evaluative adjective phrase). If the definition of adjective phrase has been expanded to include ASCRIPTIVE noun phrases[1] then our analysis is compatible. Unfortunately there is no discussion in Zelinsky-Wibbelt as to how effective their rules are when actually used in a machine translation system so we cannot make a quantitative comparison.

Murata and Nagao (1993) distinguish between GENERIC and NON-GENERIC (which is further divided into DEFINITE and INDEFINITE), using heuristics similar to rewriting rules in expert systems. They make no distinction between REFERENTIAL and ASCRIPTIVE for non-generic noun phrases. This leaves open the possibility for conflict with their rule that a noun phrase will be definite if it has been presented previously. Consider the following sentence[2]: *zō-wa honyūrui da-si, manmosu-mo honyūrui da.* 'Elephant-TOP mammal be-and mammoth-ALSO mammal be.' *Elephants are mammals and mammoths are also mammals.* This will become *Elephants are mammals and mammoths are also <u>the</u> mammals* using the rules given. Distinguishing between REFERENTIAL and ASCRIPTIVE prevents this kind of problem from occurring. We compare their results to ours quantitatively in Section 5.

3   Determination of noun phrase reference

All proper nouns are, by definition, REFERENTIAL. The algorithm used to determine the referential property of noun phrases headed by common nouns is shown in Figure 1. The algorithm presented is based on single sentences, it does not address the considerable problems of using information from outside the sentence being considered[3].

It is possible for the algorithm to be applied to the Japanese parse tree as part of the semantic analysis[4]. In **ALT-J/E,** however, the algorithm is applied after the semantic analysis has finished, during the transfer stage, because much of the semantic information is stored in the transfer dictionaries where the combination of Japanese and English makes it easy to disambiguate word senses. The overall process of translation in **ALT-J/E** is divided into seven parts. First, the system splits the Japanese text into morphemes and assigns parts of speech. Second, it parses the segmented text, often giving multiple possible interpretations.

---

[1]We feel this expanded definition is plausible, since the copula and ascriptive noun phrase combination fulfills the same semantic role as the copula and adjective phrase, that is, to ascribe a property.

[2]Examples are given with the (romanized) Japanese original, a gloss and the human translation. The examples have been simplified to exemplify points more clearly; a new translation has been made for each simplified sentence. Japanese particles are glossed as follows: TOP for *wa* which marks the topic, OBJ for *o* which marks the object and GEN for *no* which shows a genitive relation.

[3]Algorithms to use contextual information from outside the sentence are currently being implemented.

[4]For information retrieval it is obviously essential to determine the referentiality of noun phrases as part of the source language analysis.

Third, it rewrites complicated Japanese expressions into simpler ones. Fourth, **ALT-J/E** semantically evaluates the various interpretations. Fifth, syntactic and semantic criteria are used to select the best interpretation. Sixth, the selected interpretation is transferred into English. Finally, the English sentence is adjusted to give the correct inflectional forms. The algorithm described in this section has been implemented as part of the sixth stage. However, it could be implemented as part of the fifth stage.

Rules are applied in the order shown in Figure 1, with later rules over-ruling earlier ones.

The default assumption is that a noun phrase will be used to refer to some specific entity or entities in the discourse world, i.e. that it is REFERENTIAL.

There are five rules that are applied at the sentence level, which use the meanings of verbs combined with the semantic categories of nouns[5]. These can all be overridden by subsequent rules. The subjects of verbs that predicate over an entire class, and the objects of verbs which predicate EMOTIVE ACTION or EMOTIVE STATE, are GENERIC. Verbs that trigger these rules, e.g. *evolve, die out* are marked in the lexicon (Bond et al., 1993). For copulas, the subject is GENERIC if its semantic category is a descendant of the semantic category of the object, while it's complement is taken to be ASCRIPTIVE by default[6]. Finally, appositive noun phrases will be judged to be ASCRIPTIVE, as though they were the complement of a copula.

Note that these rules are only applied if the noun phrase in question is a common noun. In sentence 1, the semantic category of *meeting place* is ACTUAL PLACE, which is a child of the semantic category of *Aoi hall* PUBLIC PLACE. *Aoi hall,* however, is a proper noun so the rule is not applied.

(1)   Jap: *kaijō-wa*   *Aoi-kaikan*   *ϕ.*
        Gloss: meeting place-TOP   Aoi hall   is
        Eng: The meeting place is the Aoi Hall

The next level of rules (level 3) applies to noun phrases modified by embedded sentences. Japanese makes no phonological, morphological, or syntactic distinctions between restrictive and non-restrictive relative clauses (Kuno 1973, p. 235). This algorithm uses a simple heuristic: a noun phrase modified by a tensed embedded sentence is REFERENTIAL.

The next level of rules (level 4) is based on post-modification in the Japanese sentence. The use of some *setsubiji* 'suffixes'[7] implies that their modificant is

---

[5]The meanings of nouns are given in terms of a semantic hierarchy of 2,800 nodes. Each node is called a semantic category. Edges in the hierarchy represent IS-A relationships, so that the child of a semantic category IS-A instance of it. For example, ORGAN IS-A BODY-PART (Ogura et al. 1993).

[6]If the complement is later judged to be REFERENTIAL by a subsequent rule it is equivalent, to judging that the copula has been used equatively.

[7]*setsubiji* are a Japanese part of speech made up of suffixes that cannot stand alone, but change the meaning of the word they modify.

1. The default is REFERENTIAL

2. Sentence level rules

   (a) the subject of a verb marked in the lexicon as predicating over an entire class is GENERIC:
   manmosu-wa *zetsumetsu-shita* '<u>Mammoths</u> died out'

   (b) if the semantic category of the subject of a copula is a descendant of the semantic category of the object then the subject is GENERIC:
   manmosu-wa *dōbutsu-da* '<u>Mammoths</u> are animals'

   (c) the object of a verb which predicates EMOTIVE ACTION or EMOTIVE STATE is GENERIC:
   watashi-wa <u>*manmosu-wo*</u> *suki-da* 'I like <u>mammoths</u>'

   (d) the complement of a copula is ASCRIPTIVE:
   manmosu-wa <u>*dōbutsu-da*</u> 'Mammoths are <u>animals</u>'

   (e) appositive noun phrases are ASCRIPTIVE:
   <u>*denwagaisha-no*</u> *NTT* 'NTT, <u>a telephone company</u>'

3. Modification by embedded sentences

   (a) A noun phrase whose head is modified by a tensed relative clause is REFERENTIAL:
   <u>*kinou kita otoko*</u> 'the man who came yesterday'

4. Post-modification by *setsubiji* 'suffixes' and *joshi-sōtōgo* 'pseudo-particles'

   (a) the modificant of *muke* 'aimed at', *yō* 'for' .. .is GENERIC:
   <u>*josei-muke*</u>*-no zasshi* 'A magazine <u>for women</u>'

   (b) the modificant of *-to-iu-no-wa* 'things called' is GENERIC:
   <u>*kikai hon'yaku-to-iu-no-wa*</u> *muzukashii* '<u>Machine translation</u> is difficult'

5. Modification by demonstratives, numerals and the genitive construction *no* 'of'

   (a) A noun phrase whose head is modified by a demonstratives or numeral is REFERENTIAL:
   <u>*kono otoko*</u> 'this man', <u>*futari-no otoko*</u> 'two men'

   (b) A noun phrase whose head is modified by the genitive construction is REFERENTIAL:
   <u>*hana-no saki*</u> 'the tip of my nose'

6. A noun phrase with a 'unique' referent is REFERENTIAL:
   <u>*chikyū*</u> 'the earth'

Figure 1: Determination of noun phrase referentiality

GENERIC. For example *muke* 'aimed at' in *josei-muke-no-zasshi* 'woman aimed-at GEN magazine' *a magazine, aimed at women.* Similarly the construction *A-to-iu-no-wa* 'things called A' implies that its modificant is GENERIC. It can in fact be thought of as a pseudo-particle, the whole construction acting as a single marker which has the effect of marking its modificant as being a generic noun phrase used as the topic[8].

The next level of rules (level 5) makes a noun phrase whose head is modified by a demonstrative, numeral or the genitive construction *NP-no* 'NP's' REFER-ENTIAL. Note that only noun phrases modified by *no* judged to be genitive are REFERENTIAL. Partitive constructions such as *ōkami-no-mure* 'pack of wolf' *a pack of wolves* are not included in this judgment. The genitive construction may be translated into English in a variety of ways including a prepositional phrase headed by 'of', a possessive phrase with a clitic in the determiner position, or a possessive pronoun.

Finally (level 6), noun phrases headed by nouns that are marked in the lexicon as likely to have a unique referent, such as *chikyū* 'the earth' are assumed to be REFERENTIAL.

The algorithm presented in this section is only heuristic. Further work remains to be done to refine it. In particular: using the wa/ga distinction in conjunction with noun anaphora relations to distinguish between GENERIC and REFERENTIAL. and improving the rules at level 3 for relative clauses.

4   Using noun phrase referentiality to select articles and determine number

Knowledge of a noun phrase's referential use is essential when translating from Japanese to English, as it plays a large part in determining how a noun phrase is expressed in English. In this section we show how articles and number are generated differently for the three different referentialities in the machine translation system **ALT-J/E.** Correct generation of articles and number is important not only to express meaning accurately, but because it is one of the major factors in determining the readability of Japanese-to-English translations.

4.1   Translation of generic noun phrases

A GENERIC noun phrase (with a countable head noun) can generally be expressed in three ways (Huddleston, 1984). We call these GEN 'a', where the noun phrase is indefinite: <u>*A mammoth is a mammal:*</u> GEN 'the', where the noun phrase is definite: <u>*The mammoth is a mammal;*</u> and GEN *ϕ,* where there is no article:

---

[8]**In ALT-J/E** the entire construction (and the similar construction *A-to-iu-mono-wa* 'things called A') is rewritten during the Japanese rewriting stage into a pseudo-particle (Shirai et al., 1993), which marks its modificant as being a generic noun phrase in the ha-case (TOPIC). It is not however necessary to do this, as shown in Murata (1993), where this construction is found by matching against the Japanese dependency structure.

Mammoths are mammals. Uncountable nouns and pluralia tantum can only be expressed by GEN ϕ (eg: *Furniture is expensive)*. They cannot take GEN 'a' and they do not take GEN 'the', because then the noun phrase would normally be interpreted as having definite reference. Nouns that can be either countable or uncountable take only GEN ϕ or 'a': *Cake is delicious/ Cakes are delicious, A cake is a kind of food*. These combinations are shown in Table 1. Noun phrases that cannot be used to show GENERIC reference are marked with an asterisk (*).

Table 1: Genericness and Countability

GEN	Noun Countability Preference		
type	Countable	Both	Uncountable
'a'	a mammoth	a cake	*a furniture
'the'	the mammoth	*the cake	*the furniture
ϕ	mammoths	cake/cakes	furniture

The use of all three kinds of GENERIC noun phrases is not acceptable in some contexts, for example **a mammoth evolved*. Sometimes a noun phrase can be ambiguous, for example *I like the elephant,* where the speaker could like a particular elephant, or all elephants.

Because the use of GEN ϕ is acceptable in all contexts, **ALT-J/E** generates all GENERIC noun phrases as such, that is as bare noun phrases. The number of the noun phrase depends on the Countability preference of the noun phrase heading it and there will be no article.

4.2   Translation of referential noun phrases

The Countability and number of REFERENTIAL noun phrases can be determined with heuristics that use information from the Japanese sentence along with knowledge of English Countability stored in the lexicon. This is described in Bond et al. (1994).

According to Quirk et al. (1985:p 265), for REFERENTIAL noun phrases:

> The definite article *the* is used to mark the phrase it introduces as referring to something which can be identified uniquely in the contextual or general knowledge shared by speaker and hearer.

Whether or not a REFERENTIAL noun phrase is definite or not is determined using heuristic criteria based on whether there is enough information to uniquely identify the noun phrase's referent, such as the following:

- if the head noun is marked in the lexicon as being unique:
  the earth

- if the noun phrase is made logically unique by a modifier:
  the best price

7

- if the noun phrase's referent is restrictively described:
  the man who came to dinner, the aim of this research

- direct and indirect anaphoric reference:
  I saw a cat and a dog.   The dog chased the cat.

   As the above criteria are only meaningful for REFERENTIAL noun phrases, it is essential to determine whether the noun phrase is referential as a first step.

   When it has been determined whether a noun phrase is definite or indefinite, then articles can be generated[9]. In the final stage of processing, if there is no determiner, definite noun phrases take the definite article *the*. Indefinite countable singular noun phrases will take the indefinite article *a/an,* while indefinite countable plural and uncountable noun phrases will take the zero article ϕ. This is summarized in Table 2.

Table 2: Generation of articles for referential noun phrases.

Noun Phrase Number	Definite	Indefinite
Countable singular	*the*	*a/an*
Countable plural	*the*	ϕ
Uncountable	*the*	ϕ

4.3   Translation of ascriptive noun phrases

The countability and number of predicativeASCRIPTIVE noun phrases matches that of their subject, and the countability and number of two appositive noun phrases match each other as described in Bond et al. (1994), with the following proviso. If one element is plural and the other is a collective noun such as *group,* then they need not match. For example, *many insects, a whole swarm, . . .* as opposed to *many insects, bees I think, ...*

   ALT-J/E makes the simplifying assumption that all ASCRIPTIVE noun phrases are indefinite. Therefore, articles will be generated in the same way as for indefinite REFERENTIAL noun phrases.   Countable singular noun phrases will therefore take the indefinite article *a/an,* and countable plural and uncountable noun phrases will take the zero article ϕ.

5   Results

The processing described above has been implemented in ALT-J/E. The rules were designed using data from a specially constructed set of test sentences collected by the authors. The algorithm was evaluated on a collection of newspaper

---

[9] As well as generating definite and indefinite articles, **ALT-J/E** also generates possessive pronouns (Bond et al. 1995) and *some/any* for REFERENTIAL noun phrases when appropriate.

articles from the *Nikkei-Sangyou* newspaper by an English native speaker not connected with the development of the algorithm. The results are summarized in Table 3.

Table 3: Correct Generation of Articles and Number

	Test Sentences		Newspaper Articles	
	NPs (240)	Sentences (120)	NPs (717)	Sentences (102)
New:	94%	90%	77%	15%
Old:	70%	46%	65%	5%

New shows the results using the proposed method.
Old shows the results using the unmodified system.

We tested the system on newspaper articles, in the articles tested, there were an average of 7 noun phrases in each sentence. The articles were translated by **ALT-J/E** and the raw output examined by an English native speaker. Each noun phrase was given one of the following scores:

STRUCTURE: problem with structure or choice of translation[10]

BEST: the most appropriate article/number

ARTICLE: inappropriate article

NUMBER: inappropriate number

POSSESSIVE: inappropriate use of possessive determiner

COUNTABILITY: problem with countability

REFERENCE: problem with referential property

For the purpose of evaluating the generation of articles and number, noun phrases that were either the BEST possible translation, or that had a problem only with STRUCTURE/CHOICE OF TRANSLATION, were judged to be successful. A third-party evaluator gave the success rates as 77% for the system with the proposed method and 65% for the original system. The method of evaluation described above does not give a reproducible, absolute level of success. It does, however, successfully show the overall level of improvement/degradation, and help to iden-tify the remaining problems.

Our initial evaluation was done by the the authors, who found the success rates at the noun phrase level to be 92% for the proposed method and 76% for the system as it used to be. Nakazawa points out that this shows that the evaluation method is not reproducible (personal communication May 1995). Because the goal is to produce a translation, which is new text, there is no objective target

---

[10]This includes any major problems not connected with articles or number, such as outputing Japanese characters or spelling errors.

to compare the results with. This is a perennial problem for machine translation output. Knight and Chander (1994) in a small pilot study showed that humans could replace articles *(a/an* and *the)* in an English text in which the articles had been replaced by blanks with an accuracy of around 95%. Raw machine translation output is less coherent, than normal English text and so deciding which article is appropriate is an even harder task.

6   Discussion

In this section we discuss the remaining errors and compare the results to two other systems.

 168 of the 717 noun phrases in the machine translation of the newspaper articles had some problem. An brief analysis of the errors is given in Table 4.

Testing on the newspaper articles revealed one major heuristic that had been overlooked in the algorithm presented in section 3: some nouns when heading a construction such as 'N-*of*-NP' carry an implication that the complement NP has GENERIC reference: for example, *the applications of <u>databases</u>.* This rule will be added to the algorithm at level 5, reducing the number of errors by around 8%. Apart from this there were no major changes that needed to be made to the algorithm.

Overall, the largest sources of errors are problems with the source language analysis and dictionaries (22% each). These are not problems with the proposed algorithm but with the machine translation system as a whole. Another major source of errors is the translation of numerical expressions (12%). The processing for handling numerical expressions is currently being overhauled. The errors caused by lack of information in the dictionaries are solvable immediately, which will reduce the number of errors by around 20%.

In the generation of articles and numbers for REFERENTIAL noun phrases some of the errors can simply be solved by the addition of new rules: for example, adding rules which use the meaning of adverbs to determine number or rules using pre-head modifiers to determine definiteness. The problems of common sense deduction and indirect anaphora, however, require a large scale knowledge base and inference rules. While both are being researched at the moment, they are unlikely to be implemented soon. We estimate that the number of errors caused by insufficiencies in the generation of articles and numbers for REFERENTIAL noun phrases can be reduced at least a quarter, thus reducing the total number of errors by around 8%,.

Combining the above figures, we predict it is possible to reduce the errors by around 30%, bringing the total success rate to 84% for a window test. To go beyond this needs new processing to improve the source language analysis, the translation of numerical expressions and more use of contextual inferences.

In addition examining even this small sample of text we came up with one major addition to the algorithm for determining noun phrase reference. Therefore the algorithm needs to be tested on a wider range of texts before the rules can

Table 4: Errors in the generation of articles and number (168 noun phrases from newspaper articles)

Problem Area	Freq.	Description of error
Analysis error	22%	The Japanese noun phrase was parsed incorrectly so the rules did not trigger.
Dictionary errors	22%	The dictionary entry was incomplete.
Numerical Expressions	12%	Complicated numerical expressions are translated badly: for example *384 Kbits of networks per second* should be *a 384 Kbit/s network*
Reference	8%	There needs to be a rule to make *database* GENERIC in expressions like: *the strategic applications of databases* which is currently translated as *the strategic applications of a database*
Reference	5%	Miscellaneous errors in determining noun phrase reference.
Number	9%	In some cases rules using common sense and inference are needed to determine the number correctly: for example *sales counter* should be plural in *the sales counter of telephone companies through out the country*
Number	2%	There are no rules to deduce number from information given by adverbs: for example *prices* should be plural in *The price is 5 yen and 15 yen respectively*
Articles	7%	The rules for deciding whether a noun has been restrictively described by an embedded sentence are too coarse.
Articles	6%	There needs to be a rule for indirect anaphora. *two models* should be definite in *NTT introduced video-tel 111 and video-tel 222 in June. Two models are the first to have video receivers.*
Articles	3%	There needs to be a rule to make a noun phrase definite if its pre-head modifier restricts it sufficiently: for example *NTT will enter a video rental business*
Articles	4%	Miscellaneous errors in determining whether a noun phrase is definite or not.

be considered comprehensive. We have started testing the algorithm on a larger corpus of newspaper articles and are investigating methods for automatically learning rules.

In Murata and Nagao (1993) success rates of 68.9% for referential property and 85.6% for number were given for unknown texts of the same genre as that used in development of the rules. Their approach seems effective, although we predict the lack of a ASCRIPTIVE class will cause problems. It is impossible to directly compare our results as Murata and Nagao's testing was all carried out in Japanese by the developers, so the problems of actually generating the English and getting an impartial evaluation were not addressed. Setting these considerations aside, when we separate our results for noun phrase reference (counting as failures noun phrases with errors in article use, noun phrase reference or the use of possessive determiners), and countability and number (counting as failures noun phrases with errors in number or countability), our proposed algorithm gave success rates of 74% and 85% respectively.

Another approach is that of Knight and Chander (1994), who proposed using an automated post-editor to correct articles. Their prototype has a success rate for learning to replace articles when they have been removed from English texts of 78%. At present however the prototype cannot be used to post-edit output from a typical machine translation system as it assumes the knowledge that an article should be used in a given position, which is not normally available, and that the generation rules can function using machine translation output, which has not been shown.

7   Conclusion

This paper proposes a method that uses the information available in a Japanese sentence to identify a noun phrase as being used either GENERICALLY, REFEREN-TIALLY or ASCRIPTIVELY. This distinction is shown to be both theoretically justified and practically useful. The three way distinction in noun phrase reference is used as a base to determine a noun phrase's number and to generate appropriate articles and possessive pronouns when translating from Japanese to English. Incorporating this method into the machine translation system **ALT-J/E** helped to improve the percentage of noun phrases with correctly generated articles and number from 65% to 77%. It is shown that the proposed method can be straightforwardly extended to increase the success rate to 84%.

Several problems remain to be explored. We consider the following to of primary importance:

1. Extension of the algorithm to translate texts as coherent passages, not just as single sentences.

2. Improvement of the reproducibility of the evaluation method.

3. Investigation of the coverage of the algorithm on a wider collection of texts.

Acknowledgments

References

[BOND *et al.* (1994)] BOND, FRANCIS, KENTARO OGURA, and SATORU IKE-HARA. 1994. Countability and number in Japanese to English machine translation. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94),* 32-38.

[BOND *et al.* (1995)] -----, KENTARO OGURA, and SATORU IKEHARA. 1995. Possessive pronouns as determiners in Japanese-to-English machine translation. In *Proceedings of the 2nd Pacific Association for Computational Linguistics Conference (PACLING-95).*

[BOND *et al.* (1993)] ---------, KENTARO OGURA, SATORU IKEHARA, and SATOSHI SHIRAI. 1993. Using the meanings of verbs to select the countability of English noun phrases. In *Proceedings of the 1993 IEICE Fall Conference,* 6:61-62. IEICE.

[HUDDLESTON (1984)] HUDDLESTON, RODNEY. 1984. *Introduction to the Grammar of English.* Cambridge textbooks in linguistics. Cambridge: Cambridge University Press.

[IKEHARA *et al.* (1991)] IKEHARA, SATORU, SATOSHI SHIRAI, AKIO YOKOO, and HIROMI NAKAIWA. 1991. Toward an MT system without pre-editing - effects of new methods in ALT-J/E. In *Proceedings of MT Summit III,* 101-106.

[KNIGHT AND CHANDER (1994)] KNIGHT, KEVIN, and ISHWAR CHANDER. 1994. Automated postediting of documents. In *AAAI-94.* AAAI.

[KUNO (1973)] KUNO, SUSUMU. 1973. *The Structure of the Japanese Language.* Cambridge, Massachusetts, and London, England: MIT Press.

[LYONS (1977)] LYONS, JOHN. 1977. *Semantics,* volume 2. Cambridge: Cambridge University Press.

[MURATA (1993)] MURATA, MASAKI, 1993. Research into the determination of referential property and number of nouns using Japanese structure as a guide. Bachelor's thesis, Kyoto University, Kyoto, Japan. (in Japanese)

[MURATA AND NAGAO (1993)] ----------- , and MAKOTO NAGAO. 1993. Determi-
nation of referential property and number of nouns in Japanese sentences for
machine translation into English. In *Proceedings of the Fifth International
Conference on Theoretical and Methodological Issues in Machine Translation
(TMI-93),* 218-25.

[OGURA *et al.* (1993)] OGURA, KENTARO, AKIO YOKOO, SATOSHI SHIRAI,
AND SATORU IKEHARA. 1993. Japanese to English machine translation and
dictionaries. In *Proceedings of the 44th Congress of the International Astro-
nautical Federation,* Graz, Austria.

[QUIRK *et al.* (1985)] QUIRK, RANDOLPH, SIDNEY GREENBAUM, GEOFFREY
LEECH, and JAN SVARTVIK. 1985. *A Comprehensive Grammar of the. English
Language.* Essex: Longman.

[SHIRAI *et al.* (1993)] SHIRAI. SATOSHI, SATORU IKEHARA, and TSUKASA
KAWAOKA. 1993. Effects of automatic rewriting of source language within
a Japanese to English MT system. In *Proceedings of the. Fifth International
Conference on Theoretical and Methodological Issues in Machine Translation
(TMI-93).*

[ZELINSKY-WIBBELT (1992)] ZELINSKY-WIBBELT, CORNELIA. 1992. Exploit-
ing linguistic iconism for article selection in machine translation. In *Proceedings
of the. 14th International Conference on Computational Linguistics (COLING-
92).* 792-798.