

## Dependence on Context in case of English-Japanese Machine Translation I-10

Katsuyuki SHIBATA<sup>1)</sup>

(Received May 30, 2009)

### 英和機械翻訳における文脈依存性 I-10

柴田 勝征<sup>1)</sup>

(平成 21 年 5 月 30 日受理)

#### Abstract

This is the tenth of a series of articles on the context dependency analysis in case of English-Japanese machine translation system which we call “US system”. In this issue we examine the examples taken from Lesson 10 of four English textbooks for the first year grade Japanese junior high school students.

The main subjects to be discussed in this issue are the followings: (1) Should “their” be translated into “karera no”, “kanojotachi no” or “sorera no”? (2) Is “your” singular or plural? (revisited): (3) Is “What do you do?” singular or plural? (revisited) (4) Selection among “asobu”, “puree-suru”, “wo suru”, “wo hiku” and “wo enjiru” for verb ‘play’ (revisited), (5) Is the verb transitive or intransitive in case it doesn’t appear in the preceding sentence? (6) Selection of Japanese for “Both of them” referring to the preceding sentence and the sentence before the last ; (7) A case where the Japanese “sono” is deleted from the translation of the definite article “the”; (8) A case where “Is it?” is translated as “Oya, soudesu-ka?”; and (9) Selection between “Ee, mochiron” and “Ii desu-yo” as the Japanese for “Sure.”

Key words: machine translation, context dependence, context inherited from the preceding sentences, US system.

1. Should “their” be translated into “karera no”, “kanojotachi no” or “sorera no”?

In the fifth article <sup>[2]</sup> of this series, we examined the Japanese candidates “karera”, “kanojotachi”, “hitobito”, “sorera” for pronoun ‘they’. And in connection with that, we also explained there about the selection of

the Japanese for its objective case “them”. This time we investigate the selection of adequate Japanese for its possessive case “their” among “sorera no”, “karera no” and “kanojotachi no”.

---

<sup>1)</sup> Department of Applied Mathematics, Faculty of Science, Fukuoka University, 8-19-1 Nanakuma, Jonan-ku, Fukuoka, 814-0180, Japan

福岡大学理学部応用数学科, 〒 814-0180 福岡市城南区七隈 8-19-1

Look at the Christmas tree.  
 クリスマスツリーを見なさい。  
 Kurisumasu tsurii wo mi-nasai.

The tree has a star.  
 その木には星があります。  
 Sono ki ni-wa hoshi ga ari-masu.

The lights are very pretty.  
 明りはとてもきれいです。  
 Akari wa totemo kirei desu.

Nancy likes *their* colors very much .....(1)  
 ナンシーはたいへんそれらの（それらの明りの）  
 色が気に入りました。  
 Nanshii wa taihen *sorera no* (sorera no akari  
 no) iro ga kini-iri-mashita.

The adequate Japanese for ‘their’ in example sentence (1) above is determined by the successive application of the following two choice rules:

88; 彼女; 2; E0=their; SO<>W; J-1=3; T-1=x; v850;  
 88; u; 1; E0=their; J-1=3; J-2=1; T-2=OK; T-2<>xX; SO<>h; v945;

Rule v850 is applied when the English word in question is ‘their’ and the attributes of the subject and the object (SO) of the preceding sentence do not contain ‘W’ (= woman) and the English word preceding ‘their’ is a verb with third person-singular-present conjugate form. The last condition on the preceding verb indicates that ‘they’ corresponding to ‘their’ in question is not the subject of the sentence under translation but is most probably the subject or object of the preceding sentence. Therefore, if the attributes of the subject and the object of the preceding sentence do not contain ‘W’, female plural candidate “kanojotachi no” is abandoned.

On the other hand, rule v945 checks that the word preceding ‘their’ is a verb and that the last but two word is a noun which begins with an upper case letter not in the plural form conjugation (<>xX;). Similarly as in the case of rule v850, this indicates that ‘they’ corresponding to ‘their’ in question is not the subject of the sentence under translation but is

most probably the subject or object of the preceding sentence. Consequently, if the attributes of the subject and the object of the preceding sentence do not contain ‘h’ (= human), human plural candidates “karera no” and “kanojotachi no” are rejected.

We remark that rejection of the Japanese candidate with ‘W’ attribute by checking the nonexistence of ‘W’ among the attributes of the subject and the object of the preceding sentence corresponds to the analogous one for ‘them’ exhibited in [2]

22; 彼女; 2; E0=them; SZ<>xFX; OZ<>W; TA<>W; C897;

and also to

22; 彼女; 2; T-1<>r; TA<>W; TZ<>W; SO<>W; K630;  
 for ‘they’ presented in [4].

## 2. Is “your” singular or plural? (revisited)

In the second article<sup>[1]</sup> of this series, we explained about the following choice rule for the Japanese candidates of ‘your’, which rejects the plural interpretation of ‘your’:

88; あなた達の; 2; E0=your; SZ<>h; d642;

This rule checks the nonexistence of human attribute ‘h’ among the attributes of the subject of the preceding sentence (SZ<>h;). In such a situation, our translation system assumes that ‘you’ in question be a singular interlocutor as a default interpretation. This time we present an analogous choice rule for rejecting plural interpretation of ‘your’.

Nancy: Come here, Kenji. This is our Christmas dinner.

ナンシー「ここに来なさい、健二。これは私達のクリスマスディナーです。」

Nanshii「Koko ni ki-nasai, Kenji. Kore wa watashitachi no Kurisumasu dinaa desu」.

Kenji: Oh, it’s a wonderful turkey! Does *your* mother cook a turkey every year? .....(2)

健二「ああ、それ（クリスマスディナー）はすばらしい七面鳥です！あなたのお母さんは毎年七面鳥を料理しますか？」

Kenji “Aa, sore (Kurisumasu dinaa) wa subarashii shichimenchou desu! *Anata no* okaa-san wa mai nen shichimenchou wo ryouri-shi-masu ka?”

Nancy: Yes, she does. Does *your* mother cook one in Japan? ..... (3)

ナンシー「はい、料理します。あなたのお母さんは日本でそれを料理しますか？」

Nanshii “Hai, ryouri-shi-masu. *Anata no* okaa-san wa Nihon de sore wo ryouri-shi-masu ka?”

Kenji: No, she doesn't.

健二「いいえ、料理しません。」

Kenji “Iie, ryouri-shi-masen.”

As for the selection of the Japanese for ‘your’ in sentence (2) above, the following same rule

88; あなた方の; 2; E0=your; SZ<>h; w010;

as explained in the second article<sup>[1]</sup> of this series is applied to reject the plural interpretation ‘anatagata no’. We remark that the Japanese for the plural ‘your’ registered in our system dictionary is modified from ‘anatachi no’ to ‘anatagata no’ on the occasion of the last version up and that the corresponding choice rules are accordingly modified.

On the other hand, as for ‘your’ in example sentence (3), the choice rule above is not applied since the subject of the preceding sentence is (your) ‘mother’ which contains human ‘h’ among its attributes. Instead, the following choice rule is applied to ‘your’ in case of (3):

88; あなた方の; 2; E0=your; SZ<>FX; w015;

This rule verifies the nonexistence of attribute ‘F’ which indicates nouns connected by conjunctions (ex. Jack and Betty) as well as that of ‘X’ which means semantically plural noun (ex. people) to reject plural interpretation ‘anatagata no’.

### 3. Is “What do you do?” singular or plural? (revisited)

In the sixth article<sup>[3]</sup> of this series, we presented a choice rule which selects the plural ‘you’ interpretation of the Japanese “Anatachi wa nani wo shimasu-ka”

for “What do you do”. The selection then was based on the fact that the subject of the preceding sentence, which is estimated to correspond to ‘you’, is of the form “xxxx and I”.

This time we investigate the following example.

*What do you do* on Sundays? ..... (4)  
毎日曜日にあなたは何をしますか？

Mai nichiyou-bi ni *anata wa nani wo* shi-masu ka? .

My family enjoys Sundays.

私の家族は日曜日を楽しみます。

Watashi no kazoku wa nichiyou-bi wo tanoshimasu.

Example sentence (4) above is situated at the head of a paragraph and thus it has no preceding context information.

In our machine translation system, the following three Japanese translations are registered in the dictionary for English clause “What do you do” as an idiomatic expression.

- (i) Anata wa nani wo shimasu-ka ..... ‘you’ singular, ‘do’ present tense,
- (ii) Anata wa nani wo shite-imasu ka ..... ‘you’ singular, ‘do’ present progressive tense,
- (iii) Anatagata wa nani wo shimasu-ka ..... ‘you’ plural, ‘do’ present tense.

According to the following choice rule, the plural interpretation (iii) above is rejected since the sentence in question has no preceding context information;

XX; あなた達は; 2; E0=what do you do; SO<>xFX; x780;

Our US machine translation system adopts singular interpretation for ‘you’ as the default value when there is no special context information. The rule above checks that the attribute sets of the subject and object of the preceding sentence do not contain any of the attributes ‘x’ (noun with plural form), ‘F’ (nouns connected by a conjunction) nor ‘X’ (collective noun).

But the sentence following (4) “My family enjoys Sundays.” seems to indicate that ‘you’ in sentence (4) implies “the members of your family”. This example shows that sometimes it is necessary

to investigate the context following the sentence in question as well as the context preceding the sentence under translation in order to avoid wrong interpretations. The latest version of our US system reads the English sentence next to the sentence under translation and the system is ready to check the words and phrases in the next sentence. Regretably we have not yet formulated an adequate choice rule for sentence (4) using this function.

4. Selection among “asobu”, “puree-suru”, “wo suru”, “wo hiku” and “wo enjiru” for verb ‘play’ (revisited)

In the eighth article<sup>[4]</sup> of this series, we presented a choice rule which rejects ‘wo hiku’ among ‘asobu’, ‘puree-suru’, ‘wo suru’, ‘wo hiku’, ‘wo enjiru’ for English verb ‘play’. This time we exhibit a rule for ‘play’ which rejects ‘puree-suru’.

My father sings Christmas songs well.  
私の父は上手にクリスマスの歌を歌います。  
Watashi no chichi wa jouzu-ni Kurisumasu no uta wo utai-masu.

We sing with him.  
私達は彼と一緒に歌います。  
Watashitachi wa kare to issho-ni utai-masu.

And Kenji *plays* the piano for us. .... (5)  
そして健二は私達の為にピアノを弾きます。  
Soshite Kenji wa watashitachi no tame ni piano wo *hiki-masu*.

For ‘play’ in example (5) above, the following choice rule is applied.

33; プレーする; 2; E0=play; E1<>again; E1<>better; E2<>well; TZ<>m; VZ<>m; S810;

This rule checks that the English word next to ‘play’ doesn’t coincide with ‘again’ nor ‘better’ and that the word following the next word is not ‘well’ and further that there is no word with attribute ‘m’ (medical, sports or health) in the rear nor that the verb of the preceding sentence does not have ‘m’ attribute.

5. Is the verb transitive or intransitive in case it doesn’t appear in the preceding sentence?

In the fifth article<sup>[2]</sup> of this series, we explained about the determination of whether a given verb is used as transitive or intransitive in the English sentence under translation. We remarked there that “it is too much simply-minded to think that a transitive verb is followed by a noun phrase (object) while an intransitive verb is not. It often happens that ..... the object of a transitive verb moves sometimes before the verb or it may even be omitted. So there is no simple criterion”.

This time, we present a choice rule which judges the given verb as intransitive after verifying that the verb in question is not the same verb in the preceding sentence and that the verb in the preceding sentence didn’t appear in an interrogative sentence;

It is cold in Hirosaki.  
弘前は寒いです。  
Hirosaki wa samui desu.

Our winter vacation *begins* on December 24.  
..... (6)  
私達の冬休みは12月24日に始まります。  
Watashitachi no fuyu-yasumi wa juuni-gatsu 24 nichu ni *hajimari-masu*.

The verb ‘begin’ in sentence (6) above can apriori be used as transitive “wo hajimeru” or intransitive “hajimaru”. The following choice rule checks that the word following ‘begin’ is a preposition different from ‘ni yotte’, that the next word (‘December 24’) to the preposition (‘on’) is either a noun or an adjective with attribute ‘n’ (number), and that the word preceding ‘begin’ is a noun (‘winter vacation’) which is preceded by an article (‘Our’) beginning with an uppercase letter. All these conditions verified, it is almost clear that the verb ‘begin’ in question is intransitive. But the choice rule checks further that the Japanese translations of the verb in question differs from that of the verb of the preceding sentence (DJ0<>VR\$) and that the verb in the preceding sentence didn’t appear in an interrogative sentence; These verifications exclude the exceptional cases where the object of the transitive verb might have been omitted.

33; \*j; 1; J1=7; DJ1<>によって; J2=15; T2=n; J-1=1; J-2=8; T-2=O; DJ0<>VR\$; VZ<>?; X825;

6. Selection of Japanese for “Both of them” referring to the preceding sentence and the sentence before the last

We reviewed in the first section that the pronoun ‘they’ has four Japanese candidates ‘karera’, ‘kanojotachi’, ‘hitobito’ and ‘sorera’, and this situation consequently gives rise to the corresponding selection problem of the Japanese for the objective ‘them’. In this section, we exhibit an example where the translation system examines the subjects of the three preceding sentences in order to select an adequate Japanese for ‘both of them’. The following example is taken from an Easop’s tale.

F: Look. *The man* is riding on the donkey. (7)  
F 「見なさい。男はそのろばに乗っています。」  
F “Mi-nasai. *Otoko* wa sono roba ni notte-i-masu.”

G: *The child* is walking. .... (8)  
G 「子供は歩いています。」  
G “*Kodomo* wa aruite-i-masu.”

H: Oh, poor little boy! .... (9)  
H 「ああ、かわいそうな小さな少年！」  
H “Aa, kawai-sou-na chiisa-na shounen!”

I: Look. Both of *them* are on the donkey. ... (10)  
私 「見なさい。彼等は二人ともそのろばの上に乗っています。」  
Watashi “Mi-nasai. Karera wa hutari-tomo sono roba no ue ni notte-i-masu.”

J: It is carrying *them* on its back. .... (11)  
J 「それは彼等を背負っています。」  
J “Sore wa *karera* wo seotte-i-masu.”

The subject “Both of them” in sentence (10) above implies the union of the subject “The man” of sentence (7) and of the subject “The child” of (8). Our US system does not register the subject of an exclamative sentence as sentence (9), and therefore the variable sj[], which usually keeps the Japanese for the subject of the preceding sentence, keeps ‘kodomo’ (Japanese for ‘child’), and the variable sj2[],

which usually keeps the Japanese for the subject of the last but one sentence, keeps ‘otoko’ (Japanese for ‘man’).

The following choice rule checks that the attributes of these subjects (sz[] and sz2[]) both contain ‘h’ (human). Our translation system avoid registering the same noun in both sj[] and sj2[] for efficiency, and consequently the values of sj[] and sj2[] are different. The rule also checks that neither the object of the preceding sentence nor the object of the last but one sentence have attributes ‘h’ (human), ‘x’ (noun in the plural form) or ‘X’ (collective noun). These indicate that the subject “Both of them” of (10) means the union of those of the preceding sentence sj[] (‘kodomo’) and of the last but one sentence sj2[] (‘otoko’). So the rule selects the Japanese candidates which contain ‘kare’ (‘彼’) as adequate.

22; \*彼; 2; E0=them; E-1=of; E-2=both; E-3=.; E-4<>\*  
; DJ-5=「; SZ=h; OZ<>hxX; SZ2=h; OZ2<>hxX; J145;

Applying this rule and other choice rules successively, the translation system determines that the adequate Japanese for ‘them’ in (10) is ‘karera’.

Using this result, the Japanese for ‘them’ in (11) is selected to be ‘karera’ by the following choice rule:

22; それら; 2; E0=them; J-1=#; E-2=.; DJ-3=「; T-4=G; T-4=R; T-5=v; E-6=it; DJ-7=「; PJS=\*彼等; PJS<>\*それら; SO<>xFX; J365;

In this rule, ‘PJS’ implies “Preceding Japanese Sentence”. The rule requires that PJS contains ‘karera’ but that it doesn’t contain ‘sorera’. The rule also requires that neither the subject nor the object of the preceding sentence has attributes ‘x’, ‘F’ or ‘X’ which would imply the plurality of the noun. The conditions described in the choice rule above preceding these verify that the sentence under translation is of the form “It is ~ing .....”, that the part “~ing” is an idiomatic expression of a special type which we call “embedded type idiom” (attribute ‘R’) and that ‘them’ in question is embedded in this idiom. In example (11) above, the embedded type idiom is “carry [(pro-noun) on [possessive] back” ([ (pro-noun) wo seotte-iru), and [(pro-noun)] is ‘them’ in question in the case of (11).

7. A case where the Japanese “sono” is deleted from the translation of the definite article “the”

In Section 6 “Whether the definite article ‘the’ should be translated as ‘sono’ or to be neglected?” of the fifth article<sup>[2]</sup> of this series, we explained that the Japanese ‘sono’ for the definite article is neglected if the noun modified by that ‘the’ appears nowhere in the preceding sentence and if the noun doesn’t coincide with any of the exceptional nouns like ‘kuni’ (country), ‘basho’ (place), ‘tango’ (word), ‘mondai’ (problem), ‘daimoku’ (title) and so on. Otherwise the system usually selects ‘sono’ as the Japanese for ‘the’. Look at the following example:

L: Look. They are carrying *the donkey*. ... (12)  
 L 「見なさい。彼等は**そのろば**を持ち歩いています。」  
 L “Mi-nasai. Karera wa *sono roba* wo mochi-aruite-i-masu.”

M: They are very foolish.  
 M 「彼等はとてもばかです。」  
 M “Karera wa totemo baka desu.”

N: *The donkey* doesn’t like it. .... (13)  
 N 「**そのろば**はそれが好きではありません。」  
 N “*Sono roba* wa sore ga suki-de-wa ari-masen.”

O: Look. *The donkey* is in the water! ..... (14)  
 O 「見なさい。ろばは水の中にいます！」  
 O “Mi-nasai. *Roba* wa mizu no naka ni i-masu!”

In case of “the donkey” in example sentence (12) above, it is supposed to indicate “the donkey” in the sentence preceding (12) (which is not quoted above). So “the donkey” in (12) is translated as “sono roba” with ‘sono’ (the Japanese for ‘the’). Analogously “the donkey” in (13) is translated as “sono roba”. This is the general rule for the translation of the definite article ‘the’. But repeating ‘sono’ for many times is not considered to be a good Japanese style, and our system sometimes deletes ‘sono’ as shown in case of (14) above. The following grammar rule deletes ‘sono’ from a noun phrase once made up of ‘sono’ + noun:

2; 1; 1; 0; DJ0=その; E0=the; PJS=\*DJ0; PJS<  
>\*それ (その; DJ0<>PREOJ2\$; DJ0<>国; D

J0<>場所; DJ0<>単語; DJ0<>問題; DJ0<>題目; D0<>冊; J0#-その: sono; 8967;

This grammar rule deletes ‘sono’ from the head of a noun phrase “sono xxxxx” if the Japanese translation of the preceding sentence contains “sono xxxxx” in order to avoid repeating ‘sono’. Between sentences (12) and (13), there is a sentence which doesn’t contain “sono roba” explicitly, and thus the rule above is not applied to (12) nor (13). Therefore there remains “sono roba” untouched in (12) and in (13).

8. A case where “Is it?” is translated as “Oya, sou-desu-ka?”

If we translate English sentence “Is it?” word by word, it would become “de-aru” + “sore” + “?”. In order to translate a responding sentence like this in a conversation, it is inevitable to analyse the context.

Kumi: This book is very interesting.  
 久美「この本はとてもおもしろいです。」  
 Kumi “Kono hon wa totemo omoshiroi desu.”

Mukami: *Is it?* ..... (15)  
 Mukami 「おや、そうですか？」  
 Mukami “*Oya, sou-desu ka?*”

The following grammar rule for the generation of the Japanese translation for sentence (15) above verifies that the verb of the preceding sentence is ‘be’ (VZ = v), and thus confirms that the sentence under translation is a responding sentence to interlocutor’s statement of pattern “A is B”. In such a case the rule deletes all the Japanese translation copied from the system dictionary and rewrites it by “Oya, sou-desu ka”.

3; 32; X; 0; FT; LT; T0=O; T0=v; VZ=v; J2=?; J0#Z; J1#Z; J0#+おや、そうですか: oya, sou-desu ka; A495;

9. Selection between “Ee, mochiron” and “Ii desu yo” as the Japanese for “Sure.”

We have two Japanese translations for interjection “Sure” in our system dictionary, namely, “Ee, mochiron” and “Ii desu yo”. Of course the choice is made by analysing the preceding sentence. If the preceding

sentence is of the type “Onegai ga aru no-desu ga” (“Will you do me a favor”), the system chooses “Ii desu yo”, and otherwise “Ii desu yo” is abandoned. The following example is to reject “Ii desu yo”.

Mukami: Can I borrow it?  
 Mukami 「私はそれを借りる事が出来るでしょうか？」  
 Mukami “Watashi wa sore wo kariru koto ga dekiru deshou ka?”  
  
 Kumi: *Sure*. ..... (16)  
 久美 「ええ, もちろん。」  
 Kumi “*Ee, mochiron.*”

The following choice rule is applied to (16) above, and thus “Ii desu yo” is abandoned and as a consequence “Ee, mochiron” is accepted.

XX; いいですよ; 2; E0=\*Sure; E0<>\*; PJS<>\*です  
 が? ; x565;

## References

- [1] Katsuyuki SHIBATA: Dependence on Context in case of English-Japanese Machine Translation I-2. Fukuoka University Science Reports, vol.35 No.2. pp.85-88, 2005  
<http://www1.rsp.fukuoka-u.ac.jp/chosho/Cntxt1-2.html>
- [2] Katsuyuki SHIBATA: Dependence on Context in case of English-Japanese Machine Translation I-5. Fukuoka University Science Reports, vol.37 No.1. pp.93-103, 2007  
<http://www1.rsp.fukuoka-u.ac.jp/chosho/Cntxt1-5.html>
- [3] Katsuyuki SHIBATA: Dependence on Context in case of English-Japanese Machine Translation I-6. Fukuoka University Science Reports, vol.37 No.2. pp.77-81, 2007  
<http://www1.rsp.fukuoka-u.ac.jp/chosho/Cntxt1-6.html>
- [4] Katsuyuki SHIBATA: Dependence on Context in case of English-Japanese Machine Translation I-8. Fukuoka University Science Reports, vol.38 No.2. pp.63-68, 2008  
<http://www1.rsp.fukuoka-u.ac.jp/chosho/Cntxt1-8.html>

