

A METHOD OF TRANSLATING ENGLISH DELEXICAL STRUCTURES INTO JAPANESE

Hideki TANAKA†, Teruaki AIZAWA, Yeun-Bac KIM and Nobuko HATADA

Science and Technical Research Labs. of NHK

1-10-11, Kinuta, Setagaya, Tokyo, JAPAN

†tanakah@strl.nhk.or.jp (junet)

1 Introduction

Common verbs such as "take," "have," "make," and "give" appear frequently in English. These verbs quite often constitute verb - deverbial noun structures such as "make an address," "give an answer" and "take an approach." The verbs in the structure are almost devoid of lexical meaning but bear syntactic information such as tense, number and person; the deverbial noun carrying the lexical meaning. The verbs, in this case, are called "delexical verbs" [Collins] or "light verbs," [Live] which refers to their lexical emptiness.

In this paper, we call such verbs Delexical Verbs (DV) and a "DV + deverbial noun structure" a Delexical Structure (DS) following the examples of [Collins]. The frequency of these verbs in actual text can be seen, for example, in the COLLINS COBUILD ENGLISH DICTIONARY, where the paragraph on the entry "take" states: "*The most frequent use of take is in expressions where it does not have a very distinct meaning of its own, but where most of the meaning is in the noun that follows it...*"

We have been developing an English to Japanese machine translation system for news broadcasts since 1989 [Aizawa][Tanaka]. The precise translation of DS's in news texts is of great importance since they are quite frequent here. We counted the number of "take" + "noun" collocations (as verb + object) in 21 months' worth of AP texts using the parser of the machine translation system. "Take" collocated with 2,188 different nouns a total of 20,271 times. Of the collocating nouns, 87 deverbial nouns were found out of the 119 deverbial nouns listed in [Live], comprising about 28% (5,726) of all occurrences. This figure strongly supports the statement in the Collins Dictionary.

Failures in DS translation typically result from producing the primary sense instead of the delexical sense of the DV, which greatly dete-

riorates the quality of the translation. For example, "make an address" becomes "enzetsu wo tsukuru," which means "create an address."

There are two possible ways of translating a DS. The first is the idiosyncratic approach, listing all the DS's with their Japanese translations in a lexical system. This approach, however, suffers from several shortcomings:

- (1) The DS's are numerous and hard to list exhaustively: some DS's allow passivization and some deverbial nouns can be modified by quantifiers, adjectives and so on. This doubles and triples the number of possible DS combinations.
- (2) This direct method is unable to infer the translation of a DS undefined in a lexicon.
- (3) The use of this approach increases the number of lexical entries, making lexical management difficult.

Another approach is to synthesize the translation of a DS using the word sense of each component with syntactic and semantic rules. The attractive part of this "synthetic approach" is that it does not suffer from the problems mentioned above. The "monosemy approach" proposed in [Ruhl] can be viewed as the extreme manifestation of the synthetic approach. A recent lexical framework [Bograev] proposes to generate the word sense instead of listing them exhaustively in a lexicon, which is similar to the synthetic approach.

However, from a practical viewpoint, not all DS's can be translated by this approach as the necessary rules have not yet been factored out. We propose the use of a DS translation method based mainly on synthesis, and the employment of an idiosyncratic approach where synthesis is difficult. To do this, DS's were categorized into three groups, called type-1, type-2, and idiomatic DS. The first two groups are translated by the synthetic method and the last group is trans-

lated by an idiosyncratic approach which can hopefully be integrated into the former part as research reveals the underlying rules. This method should provide clear distinctions between idiomatic and synthesizable DS's through the use of a set of rules, which would facilitate the management of lexical systems.

The translation rules are quite simple for the following reasons:

(1) English DS's have Japanese equivalents in many cases, and some parallels can be seen between them.

(2) Many Japanese "verbal nouns" take the form of "sahen-meishi," which become verbs by simply adding "suru" to the tail.

However, some DS's require translation in a passive sense. The conditions were factored out through semantic consideration, and were integrated into the translation rules.

The rules were implemented in the machine translation system and AP news texts were translated appropriately, thus proving the feasibility of this method.

2 Delexical Structures

This section introduces some definitions and characteristics of DS's. A DS basically takes the form:

Delexical Verb + (determiner) + Deverbal Noun

(): optional

as in "give an answer," "make an array," "take an approach," "have a battle," "take advice," and "make an announcement." Verbs such as "put," "cast," "pay," "lay," and "throw" are also known to work as delexical verbs [Live]. These are commonly-used transitive verbs.

DS's have some interesting semantic characteristics. A verb can be substituted for a DS in many cases, but the substitution is one way: thus "step" for "take a step" but not "take a move" for "move" [Live]. The verb-substitutable DS's choose one of possibly several word senses of the corresponding verb except when they are metaphoric. The selection is controlled by the co-occurring DV. Thus the same deverbal noun can form DS's which represent different meanings.

make an order ==> request for supply of goods

give an order ==> command

The DS's give a greater colloquial [Konishi]

and euphemistic impression [Yamanashi] than the corresponding verb expressions and used frequently in modern English [Live].

Another feature of a DS is that it gives the impression that the event is brief:

She gave a scream. (brief)

She screamed. (not brief)

[Collins]

Some syntactic operations are possible on DS's. For example, adjectives and possessive pronouns can modify the deverbal noun, and some DS's can be used in passive forms.

In this paper, determiners are mainly considered as the modifier of the deverbal noun since they are the basic modifier of the DS's. The verbs "take," "make," "have," and "give" are studied since these are the most common DV's. As a result, we are able to obtain a DS in the following form:

DV + (det) + NV (1)

(): optional

DV: Delexical Verb (take, make, have, give)

det: a, an, the

NV: Deverbal noun

3 Categorization of DS's

The previous section mentioned that a DS chooses a single sense of the corresponding verb. An example is shown in figure 1.

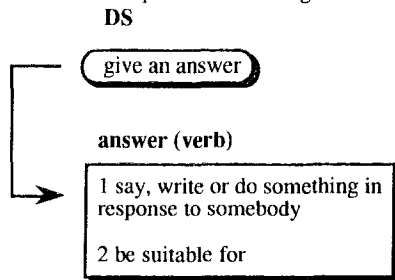


Figure 1 Word sense selection

Since a deverbal noun in a DS possesses some nominal sense of the matching verb, we can assume that a DS chooses a single sense of the deverbal noun and verbalizes it as shown in figure 2.

We term this kind of DS an ordinary DS. A DS which cannot select any sense for a deverbal noun is termed an idiomatic DS.

An ordinary DS can further be categorized into a type-1 DS or a type-2 DS. A DS which verbalizes the primary meaning of a deverbal

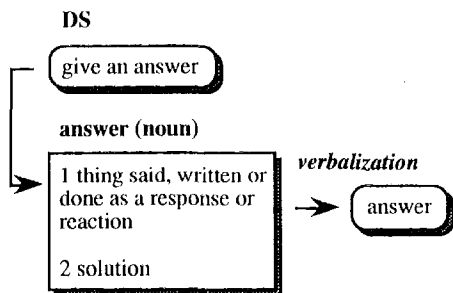


Figure 2 Word sense selection and verbalization

noun is a type-1 DS and one which verbalizes other meanings is a type-2 DS. The whole categorization is summarized below.

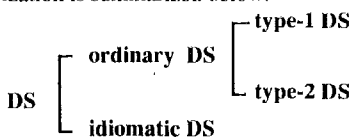


Figure 3 Categorization of DS's

4 Translation strategy

This section describes the basic strategy of DS translation. A concrete implementation will be described in section 6.

4.1 Parallels between Japanese and English

The DS categorization in section three was based on the selection of word senses for an English deverbial nouns. In many cases, we can assume parallels between English and Japanese:

- (1) The word sense of an English deverbial noun can be translated into a Japanese word sense of a deverbial noun.
- (2) The verbalization of an English word sense in a deverbial noun and the verbalization of the corresponding Japanese equivalent can represent the same meaning.

4.2 Type-1 DS

Taking advantage of these parallels, the translation of a type-1 DS can be achieved by verbalizing the core Japanese translation of the deverbial noun in a translation lexicon.

As we can see from the table in [Live], type-1 DS's are predominant. This justifies the use of type-1 as the default translation of unknown DS's.

As most of this type can be used passively, and since deverbial nouns of this type are sometimes modified by adjectives and possessive

pronouns, the translation mechanism should also allow the same degree of flexibility.

4.3 Type-2 DS

This type can be translated in approximately the same way as type-1. However, one difference is that this type verbalizes the peripheral Japanese translation in the lexicon. Since insufficient knowledge has been accrued on what makes the DS select the peripheral meaning, the collocation of DV's and deverbial nouns should be specified in a lexicon. As this type also permits some modification and allows passive transformation, it should be translated by synthetic approach.

4.4 Idiomatic DS translation

This type does not verbalize any translation of the deverbial noun and is "frozen" in that the deverbial nouns are rarely modified by adjectives and the DS does not allow passive transformation. This type of DS must be listed in a lexicon in its full form with a proper translation. The implied definition of an idiomatic DS here is (1) it does not represent any word sense of individual components and (2) it resists any form of transformation.

5 Passivization rules

Some DS's can be substituted with a matching passive verb as shown in [Live], so that "take comfort" and "take counsel" can be paraphrased as "be comforted" and "be counseled," respectively. Though the number of such cases does not seem to be large, it is important to translate them passively, otherwise the meaning of the translation is totally reversed. Most DS's which can be substituted by passive verbs should be translated in a passive sense in Japanese because of the parallels between the two languages. "Take comfort" and "take counsel" should thus be translated "nagusame rareru" and "jyogen wo sareru."

In this section, some rules for passivization are obtained by analyzing the [Live] table. Conditions for these rules will be sought in the syntactic and semantic characteristics of both the DV and the deverbial noun.

Table 1 is extracted from the table in [Live] listing the combinations requiring passive substitution.

5.1 Condition for DV

"Take" and "have" have about 10 passive verb substitutable DS's, whereas "make" and "give" have only one. This suggests that conditions depend on the contrasting characteristics

of "take," "have" and "make," "give." "Take" and "have" have a common characteristic: the action moves from the object to the subject.

With "make" and "give," on the other hand, the action moves from the subject to the object.

The characteristics of the take-have type seems to be the DV condition. Actually, we can see some hyponyms of take-have type forming passive verb substitutable DS's:

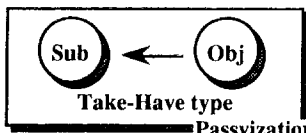
I received encouragement from my friends. --> **I was encouraged** by my friends.

Table 1 Combinations constituting passive verb substitutable DS's [Live]¹

DV's take
comfort, counsel, hint, shape, advice, encouragement, examination, offense, pleasure, satisfaction
DV's have
promise, reward, approval, disturbance, diversion, examination, excuse, impression, pleasure, reversal, satisfaction
DV's make
excuse
DV's give
excuse



Make-Give type



Take-Have type

Passivization

Figure 4 Passivization condition for DV's

5.2 Conditions for deverbal nouns

To consider the conditions for a deverbal noun, formula (1) has been elaborated on to include a subject.

Sub + DVr + (det) + NV (2)

Sub: Subject

DVr: Delexical verb requiring passive substitution (take, have)

¹ Constraints on deverbal nouns such as the determiner and the number are omitted.

As the deverbal nouns in Table 1 are derived from transitive verbs, we can picture a subject + verb + object structure for them.

X + Vt + Y (3)

X: Subject

Vt: Original Transitive Verb of the NV

Y: Object

Passive verb substitution occurs when the Sub in (2) agrees with Y in (3).

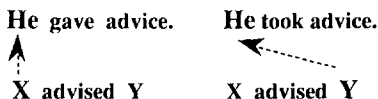


Figure 5 Agreement between Y and Sub

The conditions which make Y in (3) agree with Sub in (2) can be found in the syntactic and semantic characteristics of the verb Vt.

"Transitive" is one of the conditions mentioned previously. This is not, however, sufficient as some transitive verb derivative nouns do not require passive verb substitution.

Some examples are shown in the table 2.

Table 2 Combinations which constitute active verb substitutable DS's [Live]¹

DV's take
approach, average, bet, command, count, dare, dislike, etc.
DV's have
approach, array, battle, bet, catch, change, claim, command, debate, desire, dislike, display, dispute, doubt, dread, etc.

The deverbal nouns in Table 1 are a confined subset of Table 2. A common structure seen in the verbs in Table 1 is:

X<HUMAN> + Vt + Y<HUMAN> (4)

<>: Semantic Categorization

while many verbs in Table 2 allow both <HUMAN> and <INANIMATE> in the object slot. Furthermore, the verbs in Table 1 expect some change or action in the object <HUMAN> while the verbs in Table 2 do not. The expected change or action is up to the will of the object <HUMAN>.

X<HUMAN> + Vt + Y<HUMAN> -->

action, change (5)

Verbs of this kind generally mean "demand," "request," "order," and "permission" which have often been studied as "perlocutionary acts." In fact, we found that some nouns derived from verbs in the "perlocutionary act" group

[Yamanashi] formed passive verb substitutable DS's.

exempt

Ted *had an exemption* from military service because of his poor health.

=> Ted *was exempted* from...

permit

Did you *have permission* to enter the president's room?

=> *Were* you *permitted* to enter the president's room?

warn

Overwork apparently hastened his death, because he continued to work hard, although he had *had warning* of his illness.

=> *...had been warned...*

[Kizuka]

Formula (5) has some variations. When the subject slot is identical to the object slot, the verb becomes reflexive, thus the interpretation of (5) is: the subject expects him/herself to be affected by the action of the verb.

obligate

I *have an obligation* to submit a report to the company once a week.

=> I *obligate myself* to submit a ...²

=> I *am obliged* to submit a ...

blame

The coach *took the blame* for the poor performance of his team and resigned.

=> The coach *blamed himself* for the...

=> The coach *was blamed* for the...

[Kizuka]

The variation even allows replacement of <HUMAN> in (5) with <INANIMATE>, which should be regarded as a metaphoric variation of (5). For example, "take shape" is classified as a passive verb substitutable DS,³ thus:

My idea took shape. => My idea was shaped.

The variation of (5) can be seen in the following:

X<HUMAN, INANIMATE> shaped my idea <INANIMATE> --> action, change

In summary, the deverbal noun conditions concerning passive verb substitutable DS's are:

- (1) deverbal nouns are derived from transitive verbs,
- (2) verbs take human subjects and objects, and
- (3) objects are expected to change as the

² This is not natural but is used to reveal the underlying relationship.

³ The authors think "The idea shaped up" is also another possible paraphrase. Here we followed the table in [Live]

result of the action implied by the verb.

Condition (3) can be discussed from the viewpoint of transitivity. Since the effect on the object is strong, strong transitivity occurs [Ikegami]. However, the result of the effect is not always certain and is up to the will of the object <HUMAN>.

6 Implementation

DS translation mechanisms were implemented in an English to Japanese machine translation system [Aizawa][Tanaka] based on the discussions in sections 4 and 5.

This mechanism works after the parser has fixed the "verb + object" relationship between the DV and the deverbal noun. The new translation selection mechanism was implemented by adding some features to the existing translation selection mechanism through the use of semantic markers.

(1) Idiomatic DS

An idiomatic DS is listed in the translation lexicon as it is with its Japanese translation.

(2) Type-2 DS

Among translations of a deverbal noun, if a peripheral translation is selected by a certain DV, the type-2 marker and the co-occurring DV are specified in the marker field of the translation. The translation will be verbalized when the deverbal noun co-occurs with the DV specified in the marker field.

(3) Type-1 DS

For all deverbal nouns, the core translation will be given a type-1 marker. When a DS does not choose an idiomatic DS or a type-2 DS, the translation with a type-1 marker will be verbalized.

An example of type-1 and type-2 markers are shown in figure 6. The two translations of "exception" can be properly chosen by the description.

(4) Passivization rule

The translation of a deverbal noun which satisfies the three conditions mentioned in section 5 requires a passivization marker in addition to a type-1 or type-2 marker. When they co-occur with either "have" or "take," the translation will be verbalized in the passive voice.

(5) Verbalization

Most of the Japanese translations of English deverbal nouns are so called "sahen-meishi." Verbalization is done by simply adding "wo suru" to the tail for the active sense and "wo sareru" for the passive sense. This is realized

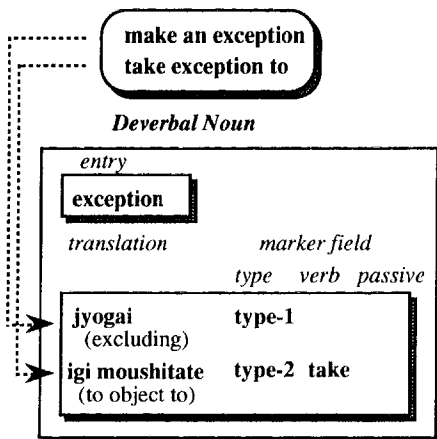


Figure 6 Lexical description for a deverbal noun

simply by rendering the translations of "make" and "give" as "wo suru," and of "have" and "take" as both "wo sareru" and "wo suru," one of which will be selected.

An example lexical description for passivization and verbalization are shown in figure 7.

"Take advice" and "give advice" are translated in passive and active voice respectively.

7 Translation Experiment

The DS's contained in the AP texts were translated using this mechanism. Two dictionaries were compiled to test the effects of the mechanism. In one of the dictionaries, necessary markers were ascribed to 302 deverbal nouns - appearing in [Live] - and 4 DV's (take, make, have, give). The other dictionary had no such modifications.

All the March 1991 AP texts which contained both the DV's and the deverbal nouns mentioned above were selected by a simple filtering of the text. The filtered text was translated using the two dictionaries and the results were compared by human assessment. Sentences not having a "verb - object" relationship between the DV and deverbal noun were rejected. The results are shown in Table 4.

Table 4 Experimental Results

DV	give	have	make	take
Improved	35	115	349	302
Equal	13	44	0	2
Made Worse	4	9	2	2
Total	52	168	351	306
Correct Rate	92%	95%	99%	99%

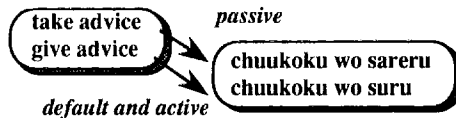
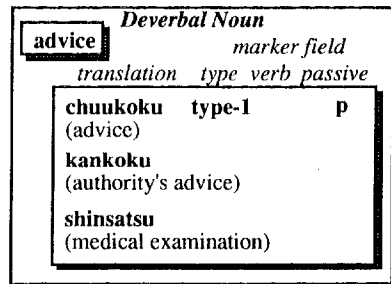
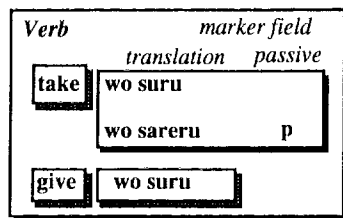


Figure 7 Lexical description for passivization and verbalization

The sentences which were translated correctly by both of the dictionaries were evaluated "Equal." The "Correct Rate," including "Equal" responses, reached nearly 100%.

Some of the results are shown below.

Upper translation is with the DS translation mechanism
Improved Examples

Case 1: But we think they'll *make a right choice* eventually.

しかし、それらが結局、正しい選択をする(sentaku wo suru)と我々は思う。
しかし、それらが結局、正しい選択を作る(sentaku wo tsukuru)と我々は思う。

Case 2: Unlike Soviets, Americans seem to have *made progress* against the stigma of alcoholism, Zubkov said. ソビエトと違ってアメリカ人はアルコール依存症の特質に対して進歩をした(shinpo wo shita) ようでであると、Zubkov が言った。ソビエトと違ってアメリカ人はアルコール依存症から特質に対する進歩を作った(shinko wo tsukutta) ようでであると、Zubkov が言った。

Case 3: After his Toyota Celica GT4 *took the lead*, he held off challengers from a trio of Lancia Delta Integrales. 彼の Toyota Celica GT4 がリードをした(riido wo shita) あと、彼は Lancia Delta Integrales のトリオから挑戦を寄せつけなかった。彼の Toyota Celica GT4 が鉛をとった(namari wo totta) あと、彼は Lancia Delta Integrales のトリオから挑戦を寄せつけなかった。

Case 4: However, they *took some encouragement* from the fact that February's decline was much smaller than the previous setbacks.

しかしながらそれは2月の下落が前の後退より大いにいっそう小さかったという、事実からいくらかの激励をされた(gekirei wo sareta)。

しかしながらそれは2月の下落が前の後退より大いにいっそう小さかったという、事実からいくらかの奨励をとった(shourei wo totta)。

Non-improved Example

Donovan "Razor" Ruddock *gives the impression* he doesn't live and breathe boxing.

Donovan Razor Ruddock は彼がしない印象をして (inshou wo shite)、そしてボクシングを呼吸する。

Donovan Razor Ruddock は彼がしない印象を与えて (inshou wo ataete)、そしてボクシングを呼吸する。

Translations without the DS translation mechanism produced the original word sense of the DV, resulting in awkward translations.

In Case 3, the DS correctly translated "take the lead" using "riido wo suru (go ahead of)" while the original system produced "namari wo toru (pick up a piece of the metal lead)." The revised mechanism thus successfully distinguished the ambiguity in the word sense of "lead." In Case 4, the passive verb substitutable DS "took some encouragement," in which the deverbial noun is modified by "some," was successfully translated passively into Japanese.

In the example which was made worse, "give an impression" was translated into "inshou wo shite," which is worse than "inshou wo ataete." This is due to the simple verbalization mechanism which attaches "suru" to all nouns backed up by heuristics. The transitive verb form of the Japanese word "inshou (impression)" is "inshou zukeru." Further elaboration of the verbalization mechanism, such as the use of a noun/verb conversion table, will solve this problem.

The quality of the translation was improved remarkably, though the translation mechanism was quite simple.

8 Conclusions

A method has been proposed for translating English delexical verb structures into Japanese.

Since this method involves rule-based synthesis, transformed and modified DS's were flexibly translated. The conditions which require a DS "passive verb substitution" were factored out and included in the translation rule. The feasibility of this method was tested by translating AP news texts, which showed a remarkable improvement in the quality of translation, with a correct rate of nearly 100%.

To achieve better results, precise verbalizations of Japanese nouns are required, though a simpler method worked well due to the ability to express many "verbal nouns" as "sahen-meishi" in Japanese.

A voice control rule was developed based on the "flow of action" of delexical verbs and the "transitivity" of deverbial nouns (the original verb). Other delexical verbs such as, "put," "cast," "receive," etc, can also be categorized by "flow of action" and can be integrated into the same framework to cover a wider range of DS translations.

The role of the other DS components, articles, possessive pronouns, and prepositions should also be explored.

The differences between "make" and "give," and "take" and "have" has been neglected in this rule. However the differences between "take" and "have" in certain contexts have been pointed out by [Wierzbicka]. It would be interesting to explore these differences in the context of translation.

[References]

- [Aizawa] T. Aizawa et al.: "A Machine Translation System for Foreign News in Satellite Broadcasting," Proc. of Coling-90 Vol. 3 pp. 308-310, (1990)
- [Boguraev] B. Boguraev and J. Pustejovsky: "Lexical Ambiguity and The Role of Knowledge Representation in Lexicon Design," Proc. of Coling-90 Vol. 2 pp. 36-41, (1990)
- [Collins] "Collins Cobuild English Grammar," Collins (1990)
- [Ikegami] Y. Ikegami: "Suru to Naru no Gengogaku," Taishuukan, (1981)
- [Kizuka] H. Kizuka: "A Dictionary of Verb - Noun Collocations," The Japan Times, (1987)
- [Konishi] Y. Konishi: "A Dictionary of English Word Grammar on Verbs," Kenkyuusha, (1980)
- [Live] A.H. Live: "The Take - Have Phrasal in English," Linguistics Vol. 95 pp. 31-50, (1973)
- [Ruhl] C. Ruhl: "On Monosemy," SUNY, (1989)
- [Tanaka] H. Tanaka: "The MT User Experience: Panel Discussion," Proc. of the MT Summit III pp. 121-126, (1991)
- [Wierzbicka] A. Wierzbicka: "Why Can You Have a Drink While *You Can't Have an Eat?," Language 58 pp. 753-799, (1982)
- [Yamanashi] M. Yamanashi: "Hatsuwa Kouji," Taishuukan, (1986)