[International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP), July 1992, Manchester]

Some Rationales and Methodologies for Example-based Approach

Makoto Nagao
Department of Electrical Engineering
Kyoto University

1 Introduction

Example-based machine translation was first proposed by the author at the International NATO Symposium on Artificial and Human Intelligence held in Lyon, France, October, 1981, and the paper was included in the book: Artificial and Human Intelligence, North-Holland, 1984 [1]. In this paper I discussed the followings in the context of comparing example based translation with ordinary machine translations which are based on phrase structure grammars or case grammars.

1) We have to learn more from the second language learning by a human and also from human translation behavior.
2) Language learners do not learn much about a grammar of a language, nor about case relations, semantic markers which are essential part of the case grammar formalism. They just learn what is given, that is, a lot of example sentences, and use them in their own sentence compositions.
3) It is quite a difficult task to write out all the possible case frames for each verb, giving exact semantic markers to each case slot for the selection of a proper noun. It is also expensive and time consuming to give semantic markers to all the nouns in a dictionary by considering every possible usages of each noun in every different context.

For non-native speakers of a language, language learning and utilization is just the imitation of native speakers' expressions either in proper or improper way. They do not learn much about grammatical structures nor about grammatical rules which the linguistic people discuss about. Ordinary people know from their common sense which examples are applicable in which situation. They usually use them by changing one or two words in each example by their similar words.

In my original paper in 1981, I showed a typical process of using parallel example phrases of Japanese and English for mechanical translation between Japanese and English. An important and difficult point here was how to get a proper example phrase for a proper part of an input sentence which was to be translated into a target language sentence. I proposed a concept of using a thesaurus which I first used in my paper [2] of sentence generation in 1965 for the semantic restriction among words in a generated sentence. Thesaurus is a tree structure representation of words by the upper-lower/wider-narrower concept relationship. Each node of a thesaurus tree corresponds to a word which is a kind of a representative of its synonymous words. A distance measure between two words was introduced on a thesaurus tree. This was calculated by the number of arcs from one word to another on a thesaurus tree. The distance of synonyms is zero.

Another problem was what part of an input sentence was to be matched by example phrases, and how a whole of an input sentence could be covered consistently by example phrases. This is still a very difficult problem, which will require a huge amount of steps, perhaps of a combinatorial explosion. Therefore this strategy should be avoided from the point of engineering system. I proposed in my original paper that the minimum structural analysis was required to escape this serious problem. A simple structural analysis gives certain phrasal structures of a sentence, to which example phrases are to be matched. This is necessary from a practical point of view.

My 1981 paper did not give impact to people in natural language processing and machine translation, but I believed the idea was good and important, and continued to tell the importance to my colleagues. Mr. Sumita who worked in our project of machine translation in 1982~ 86 was the first person who seriously considered about this, and applied the idea to the translation of a Japanese phrase "A NO B" (literal translation in English is "B of A") [3]. "A NO B" corresponds to varieties of English forms such as "B of A", "B on A", "B in A", "B for A", "AB", "A's B" and so on, according to the combination of the words A and B. According to ordinary methods semantic markers are given to A and B as S_a and S_b, and the combination of (S_a, S_b) determine the English form for "A NO B". Almost all the nouns can be A or B in "A NO B", and so semantic markers must be given to all the nouns to realize a proper selection of English forms. But the semantic markers given to nouns must also be valid in the selection of proper nouns for the case slots in case frame patterns for verbs. Therefore the attachment of semantic markers to all the nouns consistently is a very difficult problem. Actually the semantic marker system which was used in the early '80 was not accurate and sufficient. It contained about 50 semantic markers and these were not well constructed as a tree structure. The combination of (S_a, S_b) was too coarse and was not effective for the selection of proper English translation form for a Japanese expression "A NO B". Mr. Sumita collected about 600 examples of "A NO B" and their translation. He gave a best-match algorithm which calculated a kind of distance between the input and the stored examples by using thesaurus tree distance. He showed that the method was effective in the translation of "A NO B" into English. He also showed that the improvement of translation quality was quite easy because the input expressions which produced bad English translations were just added in the set of examples with their correct English translations. Example-based translation was also tried by a group of BSO in the Netherlands [4].

Features of example-based translation were discussed extensively in Sumita [3], Witkam [5], Sato [6], Sadler [7] and also in several papers of TMI-92 conference [8], [9], [10]. Therefore I do not go into the details here and would just point out the following two features.

1) Example-based approach does not create such an artificial framework as grammatical rules and semantic markers or case frames, but just refer to example sentences. Artificial framework changes itself from time to time and always is a kind of approximation to a language, while examples are always stable in a language and the accumulation of examples will cover a language more and more precisely.

2) Improvement of a system is very easy for example-based translation by increasing proper examples and their translations. Improvement for semantic marker system means the change of semantic markers which inevitably accompany the change of all the dictionary contents.

2 Example-based approach vs. semantic-marker approach

Let us discuss the similarity and the difference of example-based approach (hereafter abbreviated as EBA) and semantic-marker approach (abbreviated as SMA). EBA utilizes examples for the analysis and translation of a sentence, while SMA uses a case grammar which relies on case frame and semantic marker information. EBA uses a thesaurus tree. This tree is built by the upper-lower or broader-narrower concept relation of words. A schematic diagram of a thesaurus tree is shown in Fig. 1. Each node of this tree corresponds to a word which is a representative word of a synonymous word set. This set is shown by ( ) attached to each node in Fig. 1. When a word has two different meanings, it is located in two nodes in a thesaurus tree corresponding to its meanings. SMA is based on a set of semantic markers (abbreviated as SMs). They form sometimes just a set without any structure in it, but very often they are assembled as a tree whose parent node has a meaning broader than the meanings of daughter nodes. A schematic diagram of a semantic marker tree (abbreviated as SM tree) is shown in Fig. 2. Let us suppose that a word has one semantic marker to one word meaning. Therefore a word has as many semantic markers as the number of word meanings.

Now let us consider about the matching of an input phrase "a NO b" with example phrases in EBA. We can think of several matching algorithms, but for the purpose of comparing EBA and SMA, the following somewhat abstract description will be sufficient.

[Matching steps for EBA]

Step 1 : $a \approx A$, $b \approx B$ ($\approx$ means synonym relation). This is a complete match.

Step 2 : $a \subseteq A$, $b \subseteq B$ ($a \subset A$ means a is a lower concept of A). This is a match with a certain degree (according to the distance between a and A, and b and B in a thesaurus tree).

Step 3 : $A \subset a$, or $B \subset b$. This is a kind of mismatch (The distance between A and a is to be calculated larger for this direction).

The best match example is chosen for an input phrase, and its corresponding English translation is adopted as a kind of translation pattern for the input. When we use SMA the selection of a frame "$W(S_1)$ NO $W(S_2)$" is done by the equalness of $S_a = S_1$, and $S_b = S_2$ where S_a and S_b, are the semantic markers of a and b, and the translation is done according to the pattern associated with the frame. When we compare these two methods EBA has a property of graceful degradation, while SMA is yes/no failure. This decision procedure of SMA is not appropriate, and it is better to change the selection algorithm in the following way by assembling the semantic markers as an SM tree as shown in Fig. 2.

[Matching steps for SMA]

Step 1: $S_a \approx S_A$, $S_b \approx S_B$	This is a complete match.
Step 2: $S_a \subseteq S_A$, $S_b \subseteq S_B$	This is a reasonable match.
Step 3: $S_A \subseteq S_a$, or $S_B \subseteq S_b$	This is a mismatch.

When an SM tree is as accurate as a thesaurus, that is, when an SM tree and a thesaurus tree are essentially the same, the two approaches, EBA and SMA, are equivalent, and has the same ability. However very few semantic-marker systems have such big SM trees. In many SMAs the number of SMs are 50 ~ 60 and they are not systematically arranged into a thesaurus tree, in this case SMA becomes very coarse and is definitely inferior to EBA. That is, all the sophisticated combinations of A and B in "A NO B" can not be discriminated by specifying A and B semantically by such few SMs. Ikehara et al. [11] pointed out that an SMA by 50 SMs (which was an approximate number in Mu MT system) or by 200 SMs (which is an approximate number in the early stage of EDR dictionary project. At present EDR has 900 SMs, and is going to increase up to 4000 SMs in the coming year) is not sufficient. They constructed a very detailed case frame dictionary by establishing an SM tree of 3000 SMs. I believe that an SM tree of this size may be able to compete with a thesaurus tree system, that is, will have the same descriptive power as EBA.

It must have been a very difficult job to construct an SM tree of 3000 nodes. I guess that the only way for such a big SM system construction was to refer to a word thesaurus tree of the same preciseness and to extract a semantic key from a set of words connected to each thesaurus node. If this is true, EBA is a basis for SMA and thus superior to SMA.

A learning mechanism in EBA can be explained in the following way. It must be activated in Step 3 of EBA matching. This is the case when all the example phrases memorized so far do not match to an input expression. In this case the input expression "a NO b" or "A NO B", where A and B are representative words of the synonyms of a and b, is stored as a new example phrase. Another is a case in Step 2 when an input phrase "a NO b" matches with "A NO B" to a certain extent, but when a man judges that the match is inappropriate (that is, the English translation of the matched example "A NO B" is not a good example for an English translation of "a NO b"). In this case again it is only necessary to store the input phrase "a NO b" as a typical example phrase. The reason is the following. When an input phrase is "a NO b", this matches completely with the stored example by Step 1 of EBA after it is stored as a typical example. When an input phrase is "*a'* NO *b'*" where *a'* and *b'* are the words in the thesaurus tree shown in Fig. 3, it matches by Step 2 of EBA with the example phrase which is matched to "a NO b" at the time before "a NO b" is stored as a new example phrase. Therefore a newly stored example phrase does not do any harm to the matching of other input phrases at all. This very simple learning process is one of the best features of EBA.

Let us consider what is the process of SMA which corresponds to the above- mentioned learning process of EBA. If the SM tree is just as precise as a thesaurus tree of EBA, then "$W(S_a)$ NO $W(S_b)$ is to be stored as a new frame, where S_a and S_b are the SMs for *a* and *b* respectively. But if the SM tree is not as precise as a thesaurus tree, this new expression "*a* NO *b*" can not be represented without increasing the precision of SM tree. The increase of SMs will change a whole system completely because a new SM attachment is to be performed to all the dictionary words and to all case frames. This is almost impossible when an MT system is once constructed by an SM tree. Therefore an SM tree should be inevitably as equally precise as a thesaurus tree from the beginning.

3 A problem in choosing a proper case frame

It may be properly said that natural language analysis was done by phrase structure formalisms during 1960~1975, and was done by case grammar formalisms during 1975~1990. But as is pointed out in the previous section the case grammar formalism is too coarse to handle sophisticated language expressions and to get high quality machine translation results if the SM tree is not sufficiently precise. I feel that the present day is just the changing period from MT mechanism by using a kind of pivot expressions like case frames to a new MT framework where higher quality MT is pursued by utilizing example translation pairs.

Here we have an interesting example which shows that EBA is far better than SMA in the selection of a proper case frame for an input sentence [12]. Let us take a Japanese verb "DERU". It may correspond to "go out" as a typical meaning, but there are varieties of usages which are to be translated into different English expressions. We can use case frame information for about 500 typical Japanese verbs, which was constructed by Information Processing Promotion Association of Japan. This, abbreviated as IPAL dictionary, is said to be the best case frame information publicly available in Japan. DERU in IPAL dictionary has 32 case frames as shown in Table 1. The set of SMs in IPAL dictionary is shown in Table 2. (abbreviated as IPAL SM)

An experiment was done in the following way. First, a set of 61 test sentences which include the verb DERU was composed by a language-trained person. Particular care was taken for the test sentences to distribute well over all the different usages of DERU by consulting big Japanese dictionaries. The nouns in IPAL dictionary have IPAL SMs. These nouns are mapped into a Japanese thesaurus, BUNRUI-GOIHYO (abbreviated as BG) and all the words in BG, and therefore all the nouns in test sentences, were given IPAL SMs in this way. After this process we matched test sentences to IPAL case frames of DERU.

Case slot matching was performed roughly in the following way. Let us suppose that an input sentence has n case components and a case frame has m case slots. When n' components match between the input and a case frame by the equalness of the SMs (we regarded the lower SM in an SM tree as equal as an upper SM in the SM matching), the value:

$$const \times n' \times \sqrt{\frac{n'}{n}} \times \sqrt{\frac{n'}{m}} \Big/ \sqrt{n'}$$

is given to this matching as the matching score. Match calculation is performed for all the case frames of DERU and the best match is selected.

The result was the following.

1) There is only one best match case frame, and this is a proper case frame to be selected.	7
2) There are several best match case frames, one of which is a proper case frame to be selected.	20
3) A proper case frame was not included in the best match case frames. The matching value was lower for a proper case frame.	23
4) There is no corresponding case frame in IPAL dictionary which corresponds to an input sentence.	12

These scores indicate that the matching by SMs did not work well. The main reason is that the SMs were too coarse to distinguish every different usage of DERU.

For every case frame of IPAL dictionary there exist some typical example sentences which belong to the case frame. We performed another test which checked the best match between a test sentence and these example sentences, and as a result chose a case frame for the test sentence. The distance between words was calculated as the number of arcs to reach from a word to another in a thesaurus tree of BG. The result was the following.

1) There is only one best match case frame, and this is a proper case frame to be selected.	26

2) There are several best match case frames, one which is a proper case frame to be selected. 9

3) A proper case frame was not included in the best match case frames. The matching value was lower for a proper case frame. 15

4) There is no corresponding case frame in IPAL dictionary which corresponds to an input sentence. 12

We performed the same experiments with several other Japanese verbs. Table 3 shows the results for 19 verbs. We can see that EBA is significantly better in the selection of only one best match case frame than SMA.

These results show that the selection is far better performed by the match between example sentences and a test sentence than by the SM matching.

4 Selection of a proper translation word

By conventional MT systems a sentence is analyzed and transformed into a case frame representation, and from this expression a target language sentence is generated. In this process a target language case frame is uniquely determined from a source language case frame. When these case frames of source and target languages are the same we call the translation as a pivot method. When these case frames can be different between the source and target languages, we call the translation as transfer method.

There is a big assumption in both of these methods that the case frame correspondence between source and target languages are uniquely one to one. This is however not true. We can find out easily the examples where a same case frame representation in a source language has several target language verbs and case frames. For example, for a verb DERU there are case frames in IPAL dictionary such as

[PRO] GA [LOC] O DERU

Examples are:

The train leaves the station.
The ship clears the port.

[HUM] GA [LOC] NI DERU

Examples are:

He appeared in court.
He went to office.

These examples show that a uniquely specified case frame does not determine a unique verb in a target language. This is partly due to the coarseness of SMs, but mainly due to the freedom of language expression. Anyway, case frame representation is not enough for a proper choice of target language expressions. However, a solution for this problem is very difficult because the improvement or the increase of preciseness of SMs accompanies a heavy task to change all the dictionary contents and case frame information, and therefore SMA comes to a deadlock.

The improvement of EBA, on the other hand, can be easily achieved by adding examples in an incremental way. In the above particular case of DERU, all the alternative examples are stored as typical examples. This simple operation solves a difficult problem of the choice of proper translation words and structures.

There is, however, a problem which both EBA and SMA encounter in the stage of translation word selection. The same expression sometimes has two or more interpretations and translations. For example

DAIGAKU-O DERU $\longrightarrow$ leave the university (at five o'clock)
$\longrightarrow$ graduate from the university.

These two translations are possible because DAIGAKU can be interpreted as a place or an organization. This distinction can be done by SMA by giving different SMs to the word DAIGAKU. In the case of EBA the word DAIGAKU must be in two different places in a thesaurus tree, one under the node "place" and another under "organization".

5 A framework of EBMT

EBMT is divided into pure EBMT and hybrid EBMT [10]. Pure EBMT does not use any grammar rules, but use only example phrases to cover a whole input sentence in a consistent way. This covering problem of a sentential word string is very difficult and will raise the combinatorial explosion problem. Therefore the problem itself is theoretically interesting, but practically very inefficient. The covering problem for the syntactically analyzed sentence structure is discussed by Sato, et al. [13], and Maruyama [14]. Hybrid EBMT uses the least grammatical analysis to know which phrases modify which. This approach is inevitable from the engineering standpoint, which was recommended in my original paper in 1981. The analysis of a sentence is essentially finding a unit phrase or modifier-modifiee relations, which is essentially the dependency analysis. When a dependency structure for a sentence is obtained, example phrases which are represented as tree structures of the same nature, are to be matched to parts of this dependency tree, so that the whole structure is to be covered by example phrases. We can think of some efficient ways to match example phrases to an input sentence tree, but it is quite difficult to cover the whole tree by stored example phrases. When there remain some tree parts not covered by example phrases, the translation of these parts are to be done by the ordinary translation process of phrase structure representation. The matched portion by examples are replaced by the target language expressions (by tree structures) corresponding to these examples, and the generation of a target language sentence is performed.

By using this transfer process, translation can be realized not via a pivot representation, but by simply replacing phrasal expressions by phrasal expressions, thus escaping lots of difficulties caused by such an artificial framework as pivot representation.

6 Conclusion

Here is shown several profitable points of EBA compared to SMA. EBA can be applicable not only for machine translation but also for many other natural language processing problems. However, EBA is not almighty. Linguistic theories must be introduced into the analysis of omissions, anaphoric references, stylistics by delicate contextual information, and so on. What is discussed in this paper is that EBA gives better quality translation to SMA in the present-day machine translation frameworks. We have to utilize more linguistic information to achieve better translation quality.

References

[1] M. Nagao: A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, In A. Elithorn, ed. Artificial and Human Intelligence, Elsevier, 1984.

[2] M. Nagao: Sentence Generation by Semantic Concordance, Proc. Coling, New York, 1965.

[3] E. Sumita, et al.: Translating with Examples: A New Approach to Machine Translation, Proc. 3rd TMI-MT Conf. Texas, 1990.

[4] V. Sadler: Translating with a Simulated Bilingual Knowledge Bank (BKB), BSO/ Research, 1989.

[5] T. Witkam: MT based on Analogy, Document prepared for ATR Internal Seminar, July 26, 27, 1990.

[6] S. Sato: Example-Based Translation Approach, Proc. Int. Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP), ATR, July 1991.

[7] V. Sadler: The Textual Knowledge Bank: Design, Construction, Applications, Proc. Int. Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP), ATR, July 1991.

[8] I. J. McLean: Example-Based Machine translation Using Connectionist Matching, Proc. 4th TMI-MT Conf., Montreal, 1992.

[9] O. Furuse, H. Iida: An Example-Based Method for Transfer-Driven Machine Translation, Proc. 4th TMI-MT Conf., Montreal, 1992

[10] D. Jones: Non-Hybrid Example-Based Machine Translation Architectures, Proc. 4th TMI-MT Conf., Montreal, 1992.

[11] S. Ikehara, et al.: Semantic Analysis Dictionaries for Machine Translation, IPSJ-NLP 84-13, July 19, 1991 (in Japanese).

[12] S. Kurohashi, M. Nagao: On the Effectiveness of Example-based Approach to Semantic-marker Approach in the Selection of Japanese Case Frames, IPSJ-NLP 91-12, September 18, 1992 (forthcoming in Japanese).

[13] S. Sato, M. Nagao: Towards Memory-based Translation, Proc. Coling 90., Helsinki, 1990.

[14] H. Maruyama, H. Watanabe: Tree Cover Search Algorithm for Example-Based Translation, Proc. 4th TMI-MT Conf., Montreal, 1992.

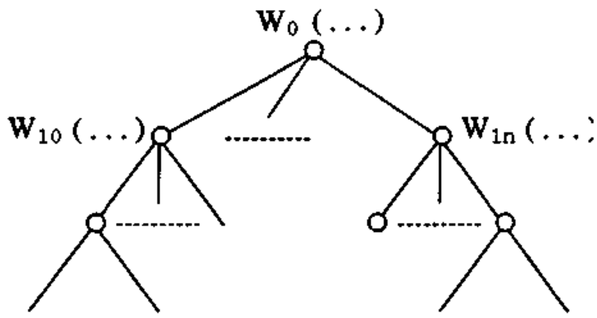


Figure 1: A thesaurus tree.

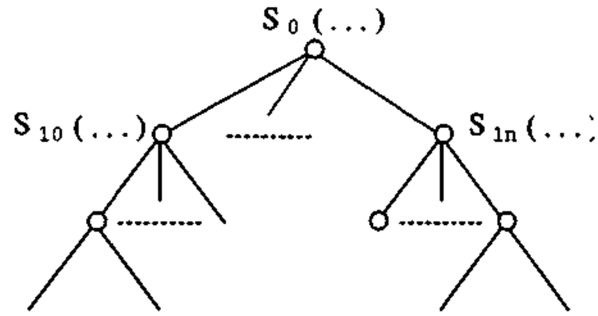
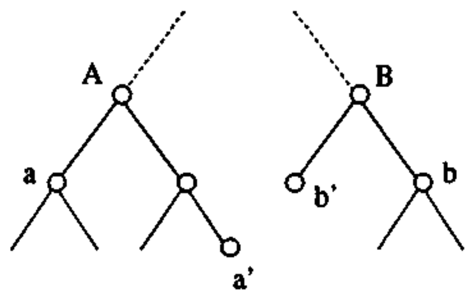


Figure 2: A semantic marker tree.



A NO B : old example phrase which matched with an input phrase "a NO b".
a' NO b' : input phrase which can match with "A NO B" after storing "a NO b" as an example phrase.

Figure 3: Explanation for the effect of storing a new example phrase.

Table 1: Case frames for DERU in IPAL dictionary (a).

ある空間の内部から外部に移動する。 — [HUM/ANI/PRO] ガ [LOC] カラ [LOC] ニ / ヘ* (彼が 部屋から 玄関に 出た。 車が 車庫から 通りに 出た。)
ある空間の内部から外部に移動する。 — [HUM/ANI/PRO] ガ [LOC] ヲ (彼が 部屋を 出る。 彼は 六時に 会社を 出た。)
内部にあった物が外部に移動する。 — [CON/PHE] ガ [LOC] カラ* [LOC] ニ / ヘ* (汗が 額から 出ている。 新芽が 出ている。)
乗り物が発進する。 — [PRO] ガ [LOC] ヲ (汽車は 駅を 出た。 船は 十四時に 港を 出ます。)
ある現象が生じる。 — [PHE] ガ [LOC/CON] カラ (火事は 隣家から 出た。 スピーカーから 音が 出る。)
身体の部分が露出する。 — [HUM] ガ [PRO/LOC] カラ* [PAR] ガ (彼女は ドレスから 肩が 出ている。)
身体の部分が突出する。 — [HUM/ANI] ガ [PAR] ガ (彼は お腹が 出ている。)
狭い所を通って、広い場所に到達する。 — [HUM/ORG/PRO] ガ [LOC] ニ / ヘ (彼は 谷間に 出た。)
ある事をするためにどこかに行く。 — [HUM] ガ [ACT] ニ [LOC] ニ / ヘ* (彼は 買物に 街へ 出た。 彼女は ヨーロッパへ 旅に 出ている。)
ある組織や所在地から去る。 — [HUM] ガ [ORG/LOC] ヲ (彼女は この三月に 大学を 出た。 彼は 家を 出て 下宿生活を始めた。)
上位の者から下位の者に何かが与えられる。 — [CON/ABS] ガ [HUM/ORG] カラ [HUM/ ORG] ニ (手当が 当局から 職員に 出る。 監督の指示が 選手に 出た。)
飲食物が与えられる。 — [TIM] ニ [PRO] ガ (夕食に ごちそうが 出た。)
ある知らせが一般に伝わる。 — [ABS] ガ [HUM/ORG/LOC] ニ (津波警報が 太平洋沿岸に 出た。)
何かが出版・発売される。 — [ABS/PRO] ガ (彼の新作が やっと 出た。)
商品が売れる。 — [CON] ガ (この品は よく 出る。)
出版物に何かが載る。 — [LIN] ニ [LIN] ガ (今朝の新聞に 彼の名前が 出ている。)

Table 1: Case frames for DERU in IPAL dictionary (b).

▼感情が身体のしぐさや表情に現れる。 — [HUM] ガ [MEN] ガ [PAR/ABS] ニ (彼は すぐに 気分が 顔に 出る。 彼女は 喜びが 表情に 出ていた。)
能力や感情などが発露される。 — [CON] ガ [MEN/CHA] ガ (彼は 勇気が 出た。 この車は スピードが 出る。)
乗り物がある速度を生じる。 — [PRO] ガ [ABS] ガ [QUA] φ (この車は スピードが ２００キロ 出る。)
ある結果が生じる。 — [PHE/ACT] デ [HUM/ABS] ガ (崖崩れで 大きな被害が 出た。)
性格や考え方などが制作物に現われる。 — [CHA/MEN] ガ [ABS] ニ (彼の性格が 作品に 出ている。 その会社のポリシーが 製品に よく 出ている。)
要求などが突き付けられる。 — [HUM/ORG] カラ [HUM/ORG] ニ [MEN] ガ (消費者から 会社に 苦情が 出た。)
ある結果が得られる。 — [ABS] ガ (何時間もの議論の末、やっと 結論が 出た。)
ゲームで、ある数字やカードなどが引き当てられる。 — [ACT] デ [ABS] ガ (サイコロで ６の目が 出た。)
何かにある特徴が備わるようになる。 — [DIV] ガ / ニ [CHA] ガ (最近の彼の作品には 重厚さが 出てきた。)
会合や試合などに参加する。 — [HUM] ガ [ACT/ABS] ニ (彼は 会議に 出た。)
ある所から何かが現れる。 — [LOC/ORG] カラ [ＮＡＴ/HUM] ガ (この地方から 文化人が 多く 出ている。 佐渡の山から 金が 出た。)
ある所に人や物が現われる。 — [LOC] ニ [HUM/ANI/ ＮＡＴ /PHE] ガ (あの廃屋には お化けが 出る。 牧場に 狼が 出た。)
紛失していた物が見つかる。 — [CON] ガ (失くしたと思っていた指輪が 出てきた。)
何かがある事に由来する。 — [LIN] ガ [ABS] カラ (この諺は 中国の古典から 出ている。 大衆の中から 出た 教え。)
ある範囲を上回る。 — [ABS] ガ [QUA/ABS] ヲ (彼女の年は 五十を 出ている。 会費は 三千円を 少し 出た。 支出が 予算を 出てしまった。)
ある態度をとる。 — [HUM/ORG] ガ [ACT] ニ (彼は 思い上がった態度に 出た。 随分 大きく 出たものだ。)

Table 2: The set of SMs in IPAL dictionary.

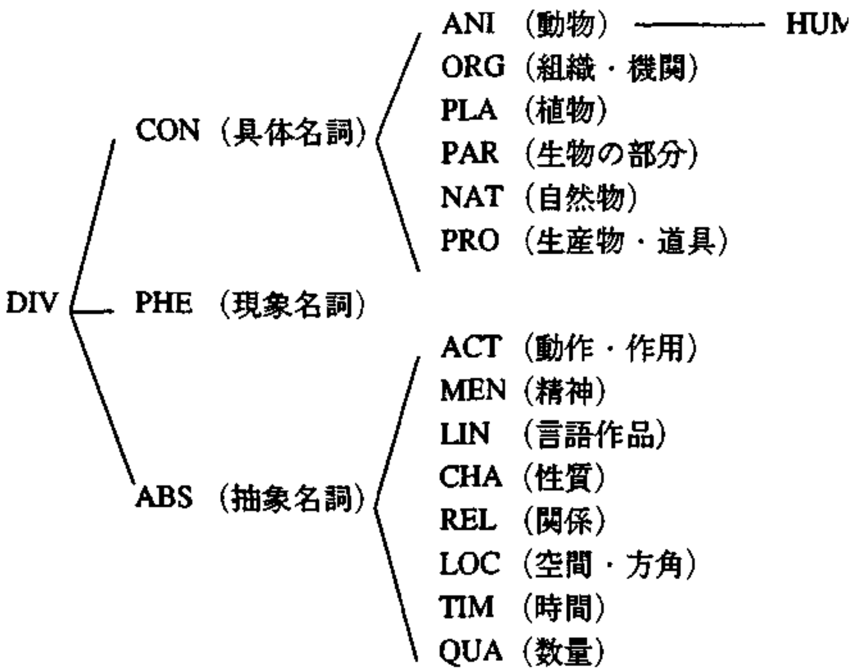


Table 3: Comparison of SMA and EBA in the selection of proper case frames for 19 verbs.

				SMA					EBA				
	i)	ii)	4)	1)	2)	3)	iii)	iv)	1)	2)	3)	iii)	iv)
(total)	-	465	52	112	195	106	27.1%	74.3%	252	58	103	61.0%	75.1%
出る	32	62	12	7	20	23	14.0%	54.0%	26	9	15	52.0%	70.0%
掛ける	28	51	8	3	22	18	7.0%	58.1%	24	3	16	55.8%	62.8%
入る	24	31	4	5	17	5	18.5%	81.5%	20	1	6	74.1%	77.8%
回る	15	37	12	8	12	5	32.0%	80.0%	15	5	5	60.0%	80.0%
おくる	15	22	1	3	12	6	14.3%	71.4%	13	3	5	61.9%	76.2%
引く	14	28	5	3	17	3	13.0%	87.0%	14	6	3	60.9%	87.0%
止まる	12	24	0	8	13	3	33.3%	87.5%	15	4	5	62.5%	79.2%
くる	12	21	2	8	5	6	42.1%	68.4%	16	1	2	84.2%	89.5%
やる	11	17	0	7	6	4	41.2%	76.5%	9	2	6	52.9%	64.7%
過ぎる	10	14	0	7	4	3	50.0%	78.6%	8	3	3	57.1%	78.6%
のる	10	25	2	5	15	3	21.7%	87.0%	13	6	4	56.5%	82.6%
抑える	9	18	0	1	12	5	5.6%	72.2%	6	4	8	33.3%	55.6%
受ける	9	20	1	9	7	3	47.4%	84.2%	15	0	4	78.9%	78.9%
突く	8	12	0	5	4	3	41.7%	75.0%	7	3	2	58.3%	83.3%
伸びる	8	20	2	3	14	1	16.7%	94.4%	9	5	4	50.0%	77.8%
合わせる	8	16	2	4	5	5	28.6%	64.3%	9	1	4	64.3%	71.4%
つかえる	8	17	1	10	1	5	62.5%	68.8%	12	1	3	75.0%	81.2%
戻る	7	17	0	11	3	3	64.7%	82.4%	11	1	5	64.7%	70.6%
追う	7	13	0	5	6	2	38.5%	84.6%	10	0	3	76.9%	76.9%

i) : number of case frames.
ii) : number of sentences.
iii) : % of 1) divided by 1) + 2) + 3)
iv) : % of 1) + 2) divided by 1) + 2) + 3)