Analogical Mapping Method and Semantic Categorization of Japanese Compound and Complex Sentence Patterns

Satoru Ikehara
Tottori University,
Tottori-city, 680-8552 Japan.
ikehara@ike.tottori-u.ac.jp

Masato Tokuhisa
Tottori University,
Tottori-city, 680-8552 Japan.
tokuhisa@ike.tottori-u.ac.jp

Jin'ichi Murakami
Tottori University,
Tottori-city, 680-8552 Japan.
murakami} @ike.tottori-u.ac.jp

Abstract

To overcome the limit of the conventional machine translation (MT) method based on *compositional semantics*, we proposed an *Analogical Mapping (AM) method* based on *Semantic Typology* and built a semantic category system for Japanese compound and complex sentences. The *AM-method* maps linguistic expressions into other expressions with the same meaning with semantic categorization (based on concepts called *Truth Items*). We also built a semantic category system composed of two sub-systems: one for classifying the meanings (222 categories) represented by the relation between two clauses and the other for classifying the meanings (740 categories) represented by clauses. These semantic codes were assigned to sentence patterns (SPs) (226,800 patterns) registered in the sentence pattern (SP) -dictionary we recently developed. We ascertained that these were useful for selecting semantically correct candidates from matched patterns in input sentences.

1. Introduction

Significant investment has been made in machine translation (MT), resulting in noteworthy achievements (Nakamura, 1983; Nagao et al., 1998; Tanaka, 1998). However, it is very difficult to develop high quality MT systems between languages belonging to very different families, such as Japanese and English.

Most practical MT systems so far have been based essentially on the *transfer method*, which is, in turn, based on *compositional semantics.* A problem with this method is that it produces translations by separating the syntactic structure from the meaning and is thus liable to lose the meaning of the source text.

Much attention has been focused on the use of *cognitive grammar* (Langacker, 1991: Lakoff, 1986), and *construction grammar* (Fillmore et al. 2003) in hopes of solving this problem. In these methods, various measures are taken to associate the meaning of constituents with the overall meaning of a sentence. However, the standards for determining the structural meaning units are undefined.

Better quality in translation can be achieved from *pattern-based MT* (Takeda, 1996) where the syntactic structure and semantics are handled together. However, this method requires immense sentence pattern (SP) dictionaries, which are difficult to develop, and so far, this method has only been used in hybrid systems where small-scale SP-dictionaries for specific fields are used to supplement a conventional *transfer method.*

Example-based MT (Nagao, 1984; Sato, 1992; Brown, 1999) might solve this problem. This method obtains translations by substituting semantically similar elements in structurally matching translation examples; hence, there is no need to prepare an SP-dictionary. However, the substitutable constituents depend on translation examples. This makes it impossible to determine them in real time. This problem could be addressed by manually tagging each example beforehand, but the resulting method would be just another *pattern-based MT.*

A *Multi-Level-Translation Method* (Ikehara et al., 1987) has been proposed and an SP-diction-

Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics, pages 181–190,

ary called *A-Japanese-Lexicon* has been developed for simple Japanese sentences (Ikehara et al., 1997) to address this problem, This dictionary includes 17,000 valence patterns of relation between verbs and case elements. Semantic use of nouns (400,000 words) is specified using *semantic attributes* (2,700 types). This dictionary has significantly improved the quality of translation of simple Japanese sentences into English.

Subsequently, a new language model that focuses on non-compositional linguistic expressions has been proposed, and a large-scale SP -dictionary (226,800 SP pairs) has recently been developed for Japanese compound and complex sentences (Ikehara, 2001).

The SPs registered in this dictionary are considered meshes for filtering out the meanings of Japanese expressions. Then, we developed an *Analogical Mapping method* (*AM-method*) that maps linguistic expressions into other expressions with the same meaning via concepts (called *Truth Items*) and built a semantic category system for the meanings of Japanese compound and complex sentences. The *AM-method* is an ideal MT method that uses an SP-dictionary.

2. *Analogical Mapping Method*

The relation between syntax and semantics has been one of the most controversial issues in language translation. In this section, we describe the *AM-method,* which is based on *Semantic Typology* (Arita 1987) and *Analogical Mapping theories* (Ichikawa, 1963) .

2.1 *Semantic Typology* of Expressions

Natural language has many "forms" of expressions, so speakers can express subtly different concepts. These forms are also used as a framework in the process of conceptualizing objects.

Jun Arita (1987) a Japanese linguist specializing in German, proposed the idea of *Semantic Typology* and *typological semantic units*. He says that *typological semantic units* exist one level below specific linguistic expressions. They are semantic structures that have been abstracted or simplified to the maximum extent practical without changing their meaning, and they are also considered as a mesh for filtering out concepts from linguistic expressions. Linguistic expressions can be analyzed and comprehended using these semantic units.

Based on this idea, translation can be considered as a process of "filtering out" cognition presented by speakers using semantic units of a source language and reproducing them in other semantic units of a target language. To produce high-quality translations, a translator must have meshes of filtering out concepts for the source language that exactly corresponds to those of the target language. If the exact grid is not available, that is, if a concept that exactly corresponds to the concept appearing in the source language is not found in the target language, it may be represented by combining more concrete or similar concepts. A large-scale SP-dictionary has recently been developed (Ikehara et al., 2006) for such a typological semantic unit.

2.2 *Analogical Mapping Theory*

Kikuya Ichikawa (1963) formulated the analogical reasoning in scientific discovery and then developed his *Analogical Mapping Theory* [*1] in *Creative Thinking*, referred to as the *Theory of Equivalent Transformation,* which states that analogical thinking lies at the core of human creativity. This theory presented a sort of model for the process of solving creative problems by assuming that different systems may have a commonality, ε, in their events or phenomena under a certain condition, C. Equation (1) shows this assumption.

$$C\ (A_{\alpha} \overset{\varepsilon}{=} B_{\beta}). \qquad (1)$$

where C is a condition, ε is a commonality,
A_{α} is an event in System α, and
B_{β} is an event in System β.

*1 In the field of artificial intelligence, human intelligence was traditionally thought to be essentially comprised of human's generic thinking rules and the reasoning (inference) ability supported by these rules. However, more attention has been paid to roles of human's analogical ability since the 1980's in AI (Suzuki, 1996).

Analogical thinking refers to the process where, given an event, A_α (source), in System α, humans conjure up an event, B_β (target), in System β that has a commonality, ε, under a condition, C.

Translation is a process that involves understanding the meaning of linguistic representations in the source language, remembering their equivalents in the target language and then selecting an appropriate expression. Thus, translation may also be based on analogical thinking, which uses eqation (1) as its assumption.

Equation (1) can be applied to any language, resulting in the following explanation. For us to translate an expression, A_α, in language α into an expression, B_β, in language β, language β must have the expression, B_β that implies a concept represented by the expression, A_α. This logic provides a basis for implementing translation between different languages based on meanings. That is, if the commonality, ε, is considered as a concept that exists in both the source and target languages, translation by semantic units is feasible.

2.3 *AM-method*

(1) Principle of *AM-method*

It is technically difficult to map the countless individual linguistic expressions of a language onto those of another language with their meanings correctly translated. However, the infinite number of expressions can be reduced to a finite number of semantic units. This method is called the *AM-method*[*1], and uses semantic units as previously discussed. Relation (2) represents the fundamental principles of the method,

$$A_\alpha \Rightarrow C(A_\alpha) \Rightarrow \varepsilon \Rightarrow C(B_\beta) \Rightarrow B_\beta, \qquad (2)$$

Where $\Rightarrow$ is a projection or mapping, ε is a *Truth Item* (a member of a *logical semantic category)* and C is a function to typify a linguistic expression as an appropriate basic semantic unit.

Relation (2) is applied to a translation if $\alpha \neq \beta$ and to rewording in the same language if $\alpha = \beta$. Although this relation represents the translation process of typological semantic patterns, i.e., the non-compositional constituents between the two languages, it can be used for processing compositional constituents.

(2) Logical Semantic Category system

As shown in Fig. 1, the semantic units of the two languages are mapped via a *Logical Semantic Category system (LSC-system)*. This system is a set of concepts called *Truth Items*.

(3) Procedure of *AM-method:*

The AM-method consists of the following steps.

Step 1: Retrieval of matched SPs.

Retrieve the SPs matched to an input sentence from the SP-dictionary.

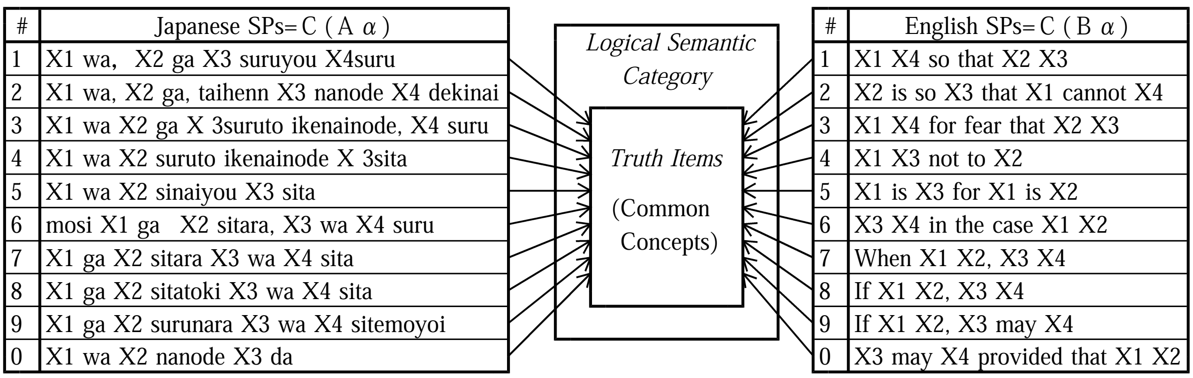


Fig. 1 Semantically equivalent mapping via *Truth Items*

*1 The translation procedure of this method is not deterministic. It is essentially different from *Example-base* or *Analogy-base* MT as well as conventional *Transfer Method*. We call it *Mapping Method* in stead of *Transfer Method*.

Step 2: SP mapping with *Truth Items.*

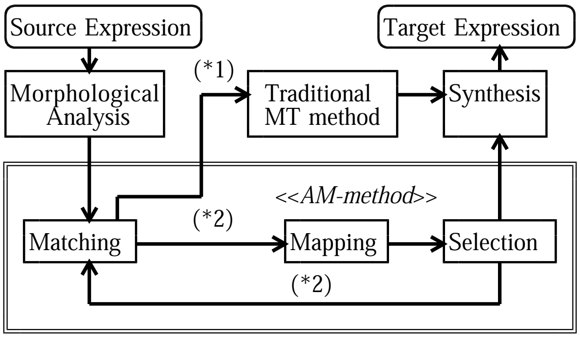
Matched SPs are mapped into semantically equivalent SPs by using *Truth Items*.

Step 3: Expression Generation by SP.

Target expressions are generated from the mapped SPs.

(4) Application to MT

The AM-method is used mainly with *non-compositional expressions*[*1] that cannot be translated by a method based on *Semantic Composition*. Because *compositional expressions* can be translated using conventional methods, both of these methods can be used together, as shown in Fig. 2. The closed loop indicates the *AM-method* is recursively used for different levels of included expressions, such as *non-compositional* clauses and phrases.



(*1) Compositional expressions
(*2) Non-compositional expressions

Fig. 2 MT System incorporated *AM-method*

3. *Logical Semantic Category system*

3.1 Target SP-dictionary

A large-scale SP-dictionary was recently developed for Japanese compound and complex sentences with two or three clauses based on the *Non-compositional language model* (Ikehara et al., 2004: 2006). Three kinds of SPs were generated from a Japanese to English parallel corpus through generalizing *compositional constituents,* as shown in the following three steps.

(a) *Word-level* SPs: Compositional independent words (nouns, verbs, etc.) were converted into *word variables*.
(b) *Phrase-level* SPs: Compositional phrases (noun phrases, verb phrases, etc.) were converted into *phrase variables*.
(c) *Clause-level* SPs: Compositional clauses were converted into *clause variables*.

The number of SPs is shown in Table 1. And their coverages are shown in Table 2. Examples of SPs will be shown later (see Fig 5)

Table 2 Coverage of SP-dictionary

Type of SP	Syntactic Coverage	Semantic Coverage
Word Level	72.1 %	55.0 %
Phrase Level	87.0 %	70.0 %
Clause Level	98.0 %	71.0 %
Total	98.5 %	79.5 %

Table 1. Number of SPs

Sentence Type	SP type	No. of Clauses	Subordinate Clause	Type of SP			
				word-level	*phrase-level*	*clause-level*	Total
Compound	Type 1	2	1 continuous clause.	53,508	36,002	17,859	109,369
	Type 2	3	2 continuous clauses	5,663	3,241	314	9,218
Complex	Type 3	2	1 adnominal clause	42,485	28,040	4,998	75,523
	Type 4	3	2 adnominal clauses	5,638	4,009	780	10,427
Mixed	Type 5	3	both types of clauses	12,510	8,146	1,524	22,280
	——	—	Total	121,904	79,438	25,475	226,817

*1 According to (Ikehara et al., 2006), *Compositional constituent* is defined as a constituent which is interchangeable with other constituents without changing *the meaning of an expression structure.* All other constituents are *Non-compositional constituents*. Based on this idea, *Compositional expression* is defined as an expression consisting of *Compositional constituents*, and *Non-compositional expression* is defined as an expression comprising one or more *Non-compositional constituents*.

We developed a *Logical Semantic Category* system to semantically classify these SPs.

3.2 Design Condition of *Truth Items*

(1) Commonality of *Truth Items*

Logical Semantic Category System consists of a set of *Truth Items. Truth Items* are defined as concepts common to the source and target languages. In general, however, perception is different from language to language even if the same object is concerned. Concepts represented by an expression are not always common to all languages.

This problem has been resolved in the SP-dictionary. Because the SPs of the source and target languages registered in the parallel corpus have a one-to-one correspondence, the concepts represented by the SP of the source the language are approximately the same as those of the target language. For this reason, assuming that the concepts represented by Japanese expressions are common to English expressions, the concepts represented by Japanese expressions are classified and used as the *Truth Items*.

(2) Granularity of Meanings

Consider to design the *Logical Semantic Category System* for Japanese compound and complex sentences, because the granularity of *Truth Items* needs to be fine enough to classify the meanings of the large-number of SPs, as shown in Table 1, it is difficult to define the meaning of an SP by using only one *Truth Item.* We defined the meaning of an SP by using multiple *Truth Items.*

Our design conditions for the system of *Truth Items* are as follows:

a) The meanings of the dependent relation between two clauses are classified by *Semantic category for clause to clause (C-to-C) relation.*
b) The meanings of clauses that compose compound and complex sentences are classified by *Semantic category for clause (C)*
c) *Truth Items* in these systems are hierarchically organized.

3.3 Construction of *LSC-system*

(1) *Semantic category for C-to-C relation*

Masuoka and Takubo (1992) classified dependent clauses into 4 types: *noun*, *adnominal*, *continuous,* and *parallel.* They analyzed the meanings of these dependent clauses and subdivided them into about 30 categories.

Table 3. *Semantic category for C-to-C relation*

L1	L2	L3	L4	Total
Compliment Clauses	noun clause	4	10	15
	interrogative clause	2	0	3
	quoted clause	2	7	10
	others	0	0	1
Noun Clauses	complemental	2	0	3
	substantial	0	2	3
	abridged modify	0	7	8
	functional	4	0	5
	"of" type	0	3	4
	others	0	0	1
Adverbial Clauses (see Fig.3)	time	2	16	19
	causality	**3**	**12**	16
	condition/concession	5	11	17
	circumstances	2	8	11
	reverse conjunction	0	9	10
	target	0	8	9
	extent	0	12	13
	premise	0	2	3
	method	0	5	6
	relation	0	7	8
	correlation	0	2	4
	descision	0	6	7
	scene	0	6	7
	authorization	0	3	4
	independent	0	6	7
	others	5	5	11
Parallel Clauses	normal parallel	6	4	11
	revers parallel	0	0	1
	others	0	0	1
others	0	0	0	1
5	29	37	151	222

(c.f.): L1-L4 represent the levels of categories. The number of each cell represents the number of categories at that level.

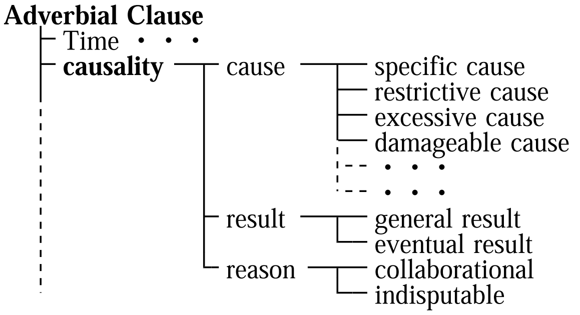


Fig. 3 Example *of Semantic Categories for C-to-C*

Referring to their study, we analyzed the meanings of the relations between clauses contained in thousands of example sentences from the view points of *temporal, spatial, logical,* and *psychological relations*, and constructed a *Semantic category for C-to-C relation*. Our *Semantic category for C-to-C* relation consists of four ranks and 222 types as shown in Table 3. Here, eamples of lower level categories under *causality* of *adverbial clause* are shown in Fig.3.

(2) *Semantic category for C*

A clause corresponds to a simple sentence, which expresses an individual event. The expression of a simple sentence is generally articulated into a propositional part and others, such as tense, aspect, and modality. The meanings of parts other than propositional part are represented by the SP structure. Subsequently, classified simple sentences by the meanings of their propositional parts and designed a *Semantic category for C.*

Simple Japanese sentences can be classified into *verb, adjective,* and *noun sentences* (Teramura, 1982) based on the types of predicates, We constructed three semantic attribute systems and used them to classify simple sentences.

Table 4. *Semantic category for C*

L1	L2	L3	L4	L5	Total
Verb Clause (see Fig.4)	**perception/emotion**	3	15	26	45
	intellectual act	6	17	20	44
	act of daily life	4	17	10	32
	act of society life	4	8	0	13
	social activity	7	23	9	40
	phenomena	9	29	11	50
	change	5	7	6	19
	movement	7	16	0	24
	objective action	4	15	8	28
Adjective clauses	nature prescription	6	34	34	75
Noun Clauses	subjective	4	26	16	47
	place	6	23	0	30
	concrete object	6	40	0	47
	abstract object	12	27	0	40
	event	4	38	98	141
	abstract relation	9	55	0	65
3	16	96	390	238	740

(c.f.:) L1-L5 represent the levels of categories. The number of each cell represents the number of categories at that level.

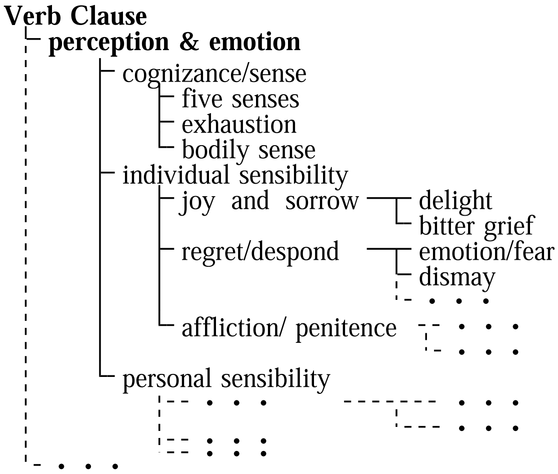


Fig. 4 Classification of perceptional or emotional Expression

The *Semantic category for C* is shown in Table 4. Here, examples of lower level categories under *perception/emotion* of *Verb Clause* are shown in Fig. 4.

Based on this classification system, we also developed a semantic verb dictionary for 6,000 words and a semantic noun dictionary for 60,000 words.

3.4 SP classification by Semantic Code

(1) SP types and semantic code

As previously stated, the SP-dictionary consists of three levels of SPs. Out of these, the *phrase level* and *clause level* SPs were obtained by further generalizing the *word level* SPs. Therefore, *Truth Items* of *word level* SPs are precursor to *phrase level* and *clause level* SPs.

Table 5. SP type and Semantic code

Semantic code	*Semantic category for C-to-C*		*Semantic category for C*		
Type of SPs	1st subordinate clause	2nd subordinate clause	1st subordinate clause	2nd subordinate clause	main clause
Type 1	○	―	○	―	○
Type 2	○	○	○	○	
Type 3	○	―	○	―	
Type 4	○	○	○	○	
Type 5	○	○	○	○	

(c.f).: Circle means applicable items

Table 5 shows the relations between the types of SPs and semantic codes. Columns 1 and 2 represent semantic codes defined by the *Se-*

mantic category for C-to-C, and columns 3 to 5 represent those of the *Semantic category for C.*

(2) Semi-automatic classification

A very large number of SPs is registered in the SP-dictionary. The meanings represented by the relation between clauses are not determined solely from conventional analyses, such as morphological and syntactic analyses. A great deal of human labor is required to assign semantic codes from the *Semantic category for C-to-C relation* to them.

This work was semi-automated by noticing that some corresponding relations can be observed between the meanings and the forms of expressions. In the case of compound sentences, we noticed the roles of conjunctive particles and functional words, and the relations between syntactic attributes of main clauses and subordinate clauses. For complex sentences, on the other hand, we noticed the relations between the type of embedded clause (inside/outside relation) and the meaning of antecedents when the complex sentence contained them.

Based on this information, we developed 90 templates for classifying the meanings of SPs. With these templates, all SPs were classified into approximately 90 groups. Semantic codes were then manually assigned to every SP in the group.

In contrast, it is not difficult to assign semantic codes of the *Semantic category for C,* because these are determined by the meanings of predicate parts of clauses. This work was automatically done by using the results of morphological and syntactic analyses and the semantic word dictionary as mentioned in 3.2 (2).

<table>
<tr><th colspan="2">Explanation</th><th>Contents registered in SP-dictionary</th></tr>
<tr><td rowspan="2">Example sentence</td><td>Japanese</td><td>kokono kikouwa watasini atteirunode rougowa kokode kurasitai
ここの 気候は わたしに 合っているので 老後は ここで 暮らしたい。</td></tr>
<tr><td>English</td><td>The climate here suits me, so I would like to live here in old age.</td></tr>
<tr><td rowspan="2">Word Level SP</td><td>Japanese</td><td>/y$1^{/tcfkN1(6810)の/kN2(15421)は}/tcfkN3(11110, 11112, 11120, 11160, 11211)に$1/cfV4(5110, 6930, 6940).teiru^rentaiので</ycN5 は>!老後は/cfADV6(9900)/fV7(3210, 3240, 6970).tai。</td></tr>
<tr><td>English</td><td>N2 ADV(N1) V4 N3^obj, so <I|N5> would like to V7^base ADV6 in old age.</td></tr>
<tr><td rowspan="2">Phrase Level SP</td><td>Japanese</td><td>/ytcfkNP1(15420, 15421)は!VP2(5110, 6930, 6940).teiru#6(.genzai|.kako)^rentaiので</ycN3 は>!VP4(3210, 3240, 6970).tai。</td></tr>
<tr><td>English</td><td>NP1 VP2#6(^present|^past), so <I|N3> would like to VP4^base.</td></tr>
<tr><td rowspan="2">Clause Level SP</td><td>Japanese</td><td>/ytcfkCL1(5110, 6930).teiru#5(.genzai|.kako)^rentaiので!CL2(3210, 6970)。</td></tr>
<tr><td>English</td><td>CL1, so CL2^past.</td></tr>
<tr><td colspan="2">Semantic code</td><td>FUb100 ／ー ‖ 4110, 5930, 5940-2210 ／ー／ 2240, 5970 ‖</td></tr>
</table>

[Explanatory notes]

Word-level SPs: ① N1, N2, N3, N5: *Noun variables.* ② V4, V7: *Verb variables.* ③ ADV6：*Adverb variable,* Here, attached bracket represents semantic attribute numbers specifying semantic constraints on a variable. ④$1^{}: constituents which canbe moved, ⑤.teiru, .tai :*tense/aspect/modaliy function,* ⑥ ^rentai, ^base, N3^ob: *word form function,* ⑦ ADV(N1) *part of speech change function,* ⑦<I|N5>: *constituent selection symbol,* ⑧ /ytcfk: Place of a constituent that may appear. (y = adonominal clause, t = constinent clause, c = case element, f = adverbial constituent, k = adonominal constituent)

Phrase-level SPs: ① NP1: *Noun phrase variable.* ② VP2, VP4: *Verb Phrase variable,* ③ VP2 #6 (^present| ^past): *selective tense description*

Clause-level SPs: ① CL1, CL2: *Clause variable*

Semantic Code: ① FUb100：*Semantic category for C-to-C* (causality-cause), ② 4110, 5930, 5940-2210: *Semantic category for C* (Subordinate clause), ③ 2240: *Semantic category for C* (Main clause)

Fig. 5 Example of semantic code assigned to SP

Fig. 5 shows an example of an SP (type 1) in which semantic codes were assigned as in Table 5..

4. Application to Selection of Matched SPs

Truth Items can be used to select SPs from matched SPs for an input sentence. When we retrieve syntactically matched SPs from the SP-dictionary, many SPs are usually obtained for one input sentence. These SP candidates include many semantically incorrect SPs. *Truth Items* are expected to be used for selecting semantically correct candidates. This section shows an example.

(1) Experimental method

We prepared two kinds of SP search programs as follows:

(a) Pattern searcher
This program compares the constituents of an input sentence and those of the patterns registered in the SP-dictionary and extracts syntactically matched SPs.

(b) Semantic searcher
This program retrieves semantically matched SPs from the SP-dictionary by comparing *Truth Items* of SPs and those of an input sentence.

The *pattern searcher* guarantees corresponding relation of constituents between an input sentence and a matched pattern, so the target expression can be generated from matched patterns. These relationships, on the other hand, are not guaranteed with the *semantic searcher*, so target expressions are not always generated from retrieved patterns. In this experiment, the SP candidates obtained by the *pattern searcher* were narrowed down with the *semantic searcher*.

(2) Experimental results

The number of SPs retrieved with each searcher for the following input sentence is shown in Table 6.

Input sentence:

watashiwa yuujinwo tatotte joukyousita
私は　友人を　頼って　上京した。
(I came to Tokyo, looking to my friend for assistance.)

Table 6. Number of retrieved SPs

Patten search program	*Word-lv.*	*Phrase lv.*	*Clause lv.*	Total
Pattern searcher	363	2,774	395	3,532
Semantic searcher	15	8	0	23
No. of common SPs	6	8	0	14

As shown in this table, many SP candidates (3,532 in total) were obtained with the *Pattern searcher*. Out of these, the top five candidates are shown below in decreasing order of the number of matched literal constituents (the order seemed to be correct).

SPs retrieved with *Pattern searcher*

△(1) /y$1/tcfkN1 を/cfV2(て|で)$1^{/ytckN3 は}/cf(V4.kako|ND4 をした)。→ N3 V(V4|ND4).past when N3 V2.past N1.
<例>話を聞いて彼は逆上した*(hanasiwo kiite karewa gyakujousita)*。"He went wild when he heard that."

×(2) /y$1/tcfkN1 を/cfV2(て|で)$1^{/ytckN3 は}/cf(V4.kako|ND4 をした)。→ N3 be.past V(V4|ND4).past to V2 N1.
<例>それを聞いて私は安心した*(sorewo kiite watasiwa ansinsita)*。"I was relieved to hear it."

×(3) /y$1/tcfkN1 を/cfV2(て|で)$1^{/ytckN3 は}/cf(V4.kako|ND4 をした)。→ V2^grn N1 V(V4|ND4).past.
<例>死体を見て彼は茫然自失した*(sitaiwomite karewabouzenjisitsusita)*。"Seeing the dead body freaked him out."

×(4) /y$1/tcfkN1 を/cfV2(て|で)$1^{/ytckN3 は}/cfV4.kako 。→ With N(V2) of N3^poss N1 N3 V4.past.
<例>手を振って彼女は立ち去った*(tewofutte kanojowa tachisatta)*。"With a wave of her hand she went away."

×(5) /y$1/tcfkNP1 を/cfV2(て|で)$1^{/ytckN3 は}!VP4.kako 。→ N3 V2.past NP1 to VP4.
<例>盲目という悪条件を克服して彼は偉大な学者になった(mekuratoiu akujoukennwo kokufukusite karewa idaina gakushani natta)。"He overcame the handicap of blindness to become a great scholar."

The meanings of the marks (○、△、×) are as follows:

○ : easy to generate a good translation
△ : semantically correct but not easy to generate a good translation
× : semantically incorrect and unusable for translation

As shown in these examples, many SPs were retrieved with the *Pattern searcher*, but very few are semantically correct.

Twenty-three SP candidates were retrieved with the *Semantic searcher*. Out of these, the top five candidates are shown below in the same order as mentioned above.

SPs retrieved with *Semantic searcher*

△(1) /y</tkN1 は>/cf いささかの/k 知辺を/cf 頼って/ytckN2 へ/cf(上っ|のぼっ|上ぼっ)た。→ <I|N1> went to N2, looking to a slight acquaintance for assistance.”
<例>いささかの知辺を頼って都へ上った *(isakakano chiennwo tayotte miyakoni nobotta)*。“I went to town, looking to a slight acquaintance for assistance.”

○(2) /y$1^{/tcfkN1 は}/tcfkN2 を/cf 頼って$1/ycf 上京した。→ N1 came to Tokyo from the country counting on N1^poss N2's help.
<例>わたしはおばを頼って上京した *(watasiha obawo tayotte joukyousita)*。“I came to Tokyo from the country counting on my aunt's help.”

○(3) /y</tkN1 は>/tcfkN2/tck 一人の/k 知人を/c 頼って/ycf 上京した。→ <I|N1> came to town, looking to an AJ(N2) friend for assistance.
<例>たった一人の知人を頼って上京した *(tattahitorino chijinnwo tayotte joukyousita)*。“I came to town, looking to an only friend for assistance.”

○(4) /y</tkN1 は>/tcfkN2 を/cf 頼って</ycfN3 は>/cf (V4.kako|ND4 をした)。→ <I|N3>N(V4|ND4) where <I|N1> could rely on <my|N1^pron^poss> N2.
<例>親類を頼って上京した *(sinnruiwo tayotte joukyou sita)*。“I went to Tokyo where I could rely on my relatives.”

×(5) /y</tkN1 は>/tcfk 雲の/k 中を/tcfkN2 に/cf 頼って/ycfV3.kako。→ <We|N1>V3.past through the clouds with the help of N2.
<例>雲の中を計器に頼って飛んだ *(kumononaka wo keikini tayotte tonda)*。“We flew through the clouds with the help of the instruments.”

From the results of these two experiments, the SP candidates can be narrowed down to 16. Out of these, the top five candidates are shown below.

Common SPs

○(1) /y$1^{/tcfkN1 は}/tcfkN2 を/cf 頼って$1/ycf 上京した。 → N1 came to Tokyo from the country counting on N1^poss N2's help.
<例>わたしはおばを頼って上京した *(watasiwa obawo tayotte joukyousita)*。“I came to Tokyo from the country counting on my aunt's help.

○(2) /y</tkN1 は>/tcfkNP2 を/cf 頼って/ycf 上京した。→<I|N1> came to town, looking to NP2 for assistance.
<例> った一人の知人を頼って上京した *(tatta hitorinochijinnwo tayotte joukyousita)*。“I came to town, looking to an only friend for assistance.”

○(3) y</tkN1 は>/tcfkN2 を/cf 頼って</ycfN3 は>/cf (V4 kako |ND4 をした)。→ <I|N3> N(V4|ND4) where <I|N1> could rely on <my|N1^pron^poss> N2.
<例>親類を頼って上京した *(sinnruiwo tayotte joukyou sita)*。“I went to Tokyo where I could rely on my relatives.

○(4) /y</tkN1 は>/tcfkNP2 を/cf 頼って/ycf 上京した。→<I|N1> came to town, looking to NP2 for assistance.
<例>たった一人の知人を頼って上京した *(tatta hitorino chijinnwo tayotte joukyousita)*。“I came to town, looking to an only friend for assistance.”

×(5) /y</tkN1 は>/tcfkN2 を /cfV3(て |で)/ycf(V4.kako|ND4 をした)。→<I|N1> N(V4|ND4) after having V3.past <my|N1^pron^poss> N2' consent by persuasion.
<例>両親を説得して上京した *(ryousinnwo settokusite joukyousita)*。“I came to Tokyo after having obtained my parents' consent by persuasion.”

As shown in this example, the number of SP candidates was narrowed down, and the ratio of semantically correct SPs significantly increased. Based on these results, *Truth Items* are promising for selecting correct SP candidates.

5. Conclusion

We proposed *an Analogical Mapping method* based on *Semantic Typology,* and we built a semantic category system for Japanese compound and complex sentences to test this method.

This system is comprised of two subsystems: one for the classifying the meaning (222 concepts) represented by the relation between two clauses and other for the classifying the meaning (740 concepts) represented by each clause.

Based on this system, semantic codes were assigned to each SP (226,800 SPs) registered in the SP-dictionary we recently developed. We ascertained that these were useful for selecting semantically correct candidates from matched SPs for input sentences.

The system of *Truth Items* developed in this study is a semantic classification system for Japanese sentences. It is an ambitious and unprecedented attempt for semantically processing of natural expressions; however it is still tentative. We are going to improve this system through various examinations and apply it to the *AM-method.* The results will be reported in the near future.

Acknowledgements

This research is supported by funding from the Japanese Government; the Core Research for Evolutional Science and Technology (CREST) awarded by the Japan Science and Technology Corporation (JST).

References

Brown, R. D. 1999. *Adding Linguistic Knowledge to a Lexical Example-Based Translation System.* TMI 99,:22-32.

Charles Fillmore, P. Kay, L. Michaelis, and I. Sag. 2005. *Construction Grammar.* Stanford Univ. Center, for Study of Language and Information Interaction.

Hideo Teramura. 1982. *Syntax and Semantics of Japanese Vol. 1.* Kurosio Publisher.

Hiroaki Suzuki. 1996. *Analogy and Thinking.* Kyoritsu Shuppan, Japan.

Hodumi Tanaka (Eds.). 1998. *Natural Language Processing: -Fundamentals and Applications-.* Iwanami Shoten, Japan.

Jun Arita. 1987. *German Lecture Course II.* Nankodo, Japan,:48-56.

Kikuya Ichikawa. 1963. *Methodology for Creative Research (extended edition).* Sanwa Shobo, Japan.

Koichi Takeda. 1996. *SP-based Machine Translation.* the 16th COLING, Vol. 2:1155-1158.

Lakoff. 1986. *Metaphors We Live By.* translated by Shoichi Watanabe and Kazuyuki Shimotani, Taishukan Shoten.

Makoto Nagao, Sadao Kurohashi, Satoshi Sato, Satoru Ikehara, and Hiroshi Nakano. 1998. *Science of Language.* Vol. 9, Language Information Processing, Iwanami Shoten.

Makoto Nagao. 1984. *A Framework of a Mechanical Translation between Japanese and English by Analogy Principle.* In A. Eithorn and R. Barneji (Eds.). *Artificial and Human Intelligence.* North-Holland, :173-180.

Ronald W. Langacker. 1991. *Foundations of Cognitive Grammar.* Vol. 1, Vol. 2, Stanford University Press.

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, and Yoshihiko Hayashi. 1987. *Speaker's cognition and a Multi-Level-Translating Method based on it.* Journal of the Information Processing Society of Japan, 28(12): 1269-1279.

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Ooyama, and Yoshihiko Hayashi. 1997 *A Japanese Lexicon.* Iwanami Shoten, Japan.

Satoru Ikehara. 2001. *Meaning Comprehension Using Semantic SPs in a Large-Scale Knowledge-Base.* Proceedings of the PACLING'01, :26-35.

Satoru Ikehara, Masato Tokuhisa, Jin'ichi Murakami, Masashi Saraki, Masahiro Miyazaki, and Naoshi Ikeda. 2006. *Pattern Dictionary Development based on Non-Compositional Language Model for Japanese Compound and Complex Sentences.* ICCPOL-06, Springer LNAI: 509-519.

Satoru Ikehara, Satsuki Abe, Masato Tokuhisa, and Jin'ichi Murakami. 2004. *Japanese to English Sentence Pattern Generations for Semantically Non-Linear Complex Sentences.* Natural Language Processing, 11(3) :70-95.

Satoshi Sato. 1992. *An example-based translation and system.* COLING-91,:1259-1263.

Takashi Masuoka and Yukinori Takubo. 1992. *Fundamental Japanese Grammar.* Kurosio Publisher, Japan.

Yasuo Nakamura. 1983. *How far can we go in translation?* The Japan Times.