A MWE-based Framework for Describing Non-propositional Content of Sentences*

Toshifumi TANABE**, Kenji YOSHIMURA**,
Yasuo KOYAMA***, and Kosho SHUDO**

Multiword expressions (MWEs) play a critical role in indicating Non-propositional Contents (NPCs) of natural Japanese sentences. We report in this paper that the proper employment of MWEs concerned enables us to put forth a general framework, which is based on a multiple nesting of semantic operations, for the processing of non-inferential NPCs of natural Japanese sentences. Our framework is characterized by its broad syntactic and semantic coverage, enabling us to deal with multiply composite modalities and their semantic/pragmatic similarity. Also, we show that the relationship between *indirect* (Searle, 1975) and *direct* speech, and the equations peculiar to modal logic or its family (Mally, 1926; Prior, 1967) are treated as similarity rules between NPCs within our framework.

Key Words: MWE, Non-Propositional Content, Illocutionary Act, Modality, Similarity

1 Introduction

While proper treatment of the Propositional Content (PC) of a sentence is undoubtedly important in natural language processing (NLP) and speech understanding (SU), the Non-propositional Content (NPC) also plays a critical role in tasks such as discourse understanding, dialogue modeling and detecting speaker's intension. We refer generically to the information which is provided by auxiliaries, adverbs, ending-particles or specific predicative forms in Japanese sentences as NPC. It is concerned with notions such as polarity, tense, aspect, voice, modality, and illocutionary act, which incorporate temporal, contingent, subjective, epistemic or attitudinal information into the PC. Though the inferential NPC e.g., *implicature* (Grice, 1975), *explicature* (Sperber and Wilson, 1995), etc., has been widely discussed in semantics or pragmatics, it lies beyond the state-of-the-art technology of NLP. Moreover, with the non-inferential NPCs, no systematic attempt to connect linguistic forms with semantic functions has been reported in NLP community. In this paper, we present a framework for the treatment of NPC of a sentence on the basis of the extensive, proper employment of multiword expressions (MWEs) indicating the NPCs in Japanese. In Japanese, which is a so-called SOV language, NPCs are typically indicated in the V-final position by auxiliaries, ending-particles and their various alternative multiword expressions. We have extensively extracted them from large-scale Japanese linguistic data. We refer to these, including auxiliaries and ending-particles, as NPC indicators (NPCIs). The number of NPCIs amounts to 1,500, whereas that of auxiliaries and ending-particles is about 50 which is apparently insufficient for practial NLP tasks. Multiword expressions play a crucial role in indicating NPCs of natural Japanese sentences.

Our model leads to dealing not only with some of *illocutionary acts* (Austin, 1962) but also with the logical operations peculiar to the family of modal logic, i.e., deontic (Mally, 1926) and temporal logic (Prior, 1967).

We also present, in this paper, the idea of the

* Received November 30, 2004
** Department of Electronics Engineering and Computer Science
*** Corporate R & D, Seiko Epson Corporation

similarity among NPCs within our framework. This is essential for text retrieval, paraphrasing, document summarization, example-based MT, etc. Some of the *indirect speech acts* (Searle, 1975) and axioms proper to the family of modal logic are treated formally in the similarity paradigm.

In Section 2, we introduce an overview of our ongoing MWE resource development for general Japanese language processing. In Section 3, we present a framework for the treatment of NPC. Sets of primitive functions to compose NPC and their indicators are explained in Section 4. In Section 5, first, the relationship between Japanese syntax and the framework, second, a statistical data concerned with the occurrence of multiword NPC indicators, and last, an experimental system for identifying NPCs of Japanese sentences are introduced. In Section 6, we discuss two application tasks of the framework: one is for machine translation, the other is for information retrieval, in particular, for the formalization of similarity between non-inferential, non-propositional contents of natural sentences. In Section 7, we present conclusions and comment on future work.

2. Background MWE-resource

The authors have been concerned with how to select atomic expressions of the sentence construction in NLP which is based on the semantic compositionality. Morphosyntactically, this problem is also serious for the processing of agglutinative, space-free languages like Japanese. Our research on this subject started in '70s by extracting manually multiword expressions as *idiomatic expressions* (Shudo et al., 1988) or *fixed-collocations* (Shudo, 1989) from large-scale Japanese linguistic data in the general domain. We estimate that the amount of data examined is 200,000 sentences.

In this Section, we present an overview of our ongoing development of Japanese MWE resources.

We have extracted multiword expressions which have at least one of the following three features;

f_1 : idiomaticity (semantic non-decomposability),
f_2 : lexical rigidity (non-separability),
f_3 : statistical boundness.

The expression which causes the difficulty in composing its overall meaning from normal meanings of component words has f_1.[(1)] f_2 includes the feature not to allow other words to cut in between the component words. The expression whose components are bound each other with high conditional probability has f_3. Each multiword expression selected as a MWE was endowed with a binary-valued basic triplet (f_1 f_2 f_3). For example, an idiomatic, separable and not-statistically-bound expression, "骨・を・折る *hone・wo・oru*" '*make an effort* (lit. *break a bone*)' is marked with (100) and a compositional, separable and statistically-bound expression, "ぐっすり・眠る *gussuri・nemuru*" '*sleep soundly*', with (001). A dot '・' denotes a conventional word-boundary and every MWE is suffixed with its triplet, hereafter.

Fixed expressions, decomposable idioms, institutionalized phrases, syntactically-idiomatic phrases and light verb constructions discussed in (Sag et al., 2002) and proverbs might correspond roughly to the triplets, (110), (100), (001), (0x1), (0x1) and (111), respectively. Here, x denotes either 1 or 0.

MWEs, whose number amounts to 64,800 at present, are classified by their overall, grammatical functions as follows. Examples and the current number of expressions are also given in the following.

Functional MWEs:

relation-indicator (RI) ＜1,000＞: "に・つい・て (110) *ni・tui・te*" '*about* (lit. *in touch with*)'; "に・よっ・て(110) *ni・yot・te*" '*by* (lit. *depending on*)'; "と・とも・に(110) *to・tomo・ni*" '*with* (lit. *accompanied with*)'; "に・おける(110) *ni・okeru*" '*in*' ,'*on* (lit. *placed in*)'; etc.

NPCI＜1,500＞: See Section 4.

Conceptual MWEs:

nominal ＜10,000＞: "赤・の・他人 (110) *aka・no・tanin*" '*complete stranger* (lit. *red stranger*)'; "鶴・の・一声 (110) *turu・no・hitokoe*" '*the voice of authority* (lit. *one note of crane*)'; etc.

verbal-nominal ＜1,700＞: "もらい・泣き (110) *morai・naki*" '*weeping in sympathy* (lit. *received crying*)'; "ラッパ・飲み (110) *rappa・nomi*" '*drinking direct from the bottle* (lit. *trumpet drink*)'; etc.

verbal ＜34,000＞: “かみ・締める (110) *kami・simeru*” ‘*chew well* (lit. *bite and fasten*)’; “煮・詰める (110) *ni・tumeru*” ‘*boil down* (lit. *boil and pack in*)’; etc.

adjectival ＜4,300＞: “怒り・っぽい (010) *okorip・poi*” ‘*irritable* (lit. *anger-ish*)’; “注意・深い (110) *chuui・bukai*” ‘*careful* (lit. *deep in caution*)’; etc.

adjectival-nominal＜2,000＞: “一巻・の・終わり (110) *ikkan・no・owari*” ‘*the very end* (lit. *the end of a roll*)’; “筋書き・通り (010) *sujigaki・doori*” ‘*as just planned* (lit. *just as a plot*)’; etc.

adverbial＜5,200＞: “悪く・する・と (110) *waruku・suru・to*” ‘*if the worst happens* (lit. *if it worsens*)’; “うっとり・と (010) *uttori・to*” ‘*abstractedly*’; etc.

adnominal＜2,600＞: “他愛・の・無い (101) *taai・no・nai*” ‘*inconsiderable* (lit. *with no altruism*)’; “断固・たる (010) *danko・taru*” ‘*firm*’; etc.

connective ＜300＞: “その・結果 (110) *sono・kekka*” ‘*consequently* (lit. *the result*)’; “それ・は・さて・おき (111) *sore・ha・sate・oki*” ‘*by the way* (lit. *setting it aside*)’; etc.

proverb-sentential＜1,300＞: “急が・ば・回れ (111) *isoga・ba・maware*” ‘*Make haste slowly.* (lit. *go round if it is in a hurry.*)’; “春眠・暁・を・覚え・ず (111) *shunmin・akatuki・wo・oboe・zu*” ‘*In spring one sleeps a sleep that knows no dawn.*’; etc.

proverb-sentential-incomplete＜900＞: “病・は・気・から (110) *yamai・ha・ki・kara*” ‘*Fancy may kill or more.* (lit. *Illness is brought from one's feeling.*)’; “馬・の・耳・に・念仏 (111) *uma・no・mimi・ni・nenbutu*” ‘*A nod is as good as a wink to blind horse.* (lit. *buddhist's invocation to the ear of a horse*)’; etc.

A set of relation indicators (RI) listed above is an extension of that of case-particles and connective-particles. Nominals, each of which is marked with a triplet (11x), include neither most of so-called compound nouns, usually marked with (00x), nor proper nouns, since their numbers amount to quite large[2]. In the lexicon, besides the regular items such as for synonym, usage (contextual structures required), etc., a MWE is endowed with a partly multi-valued 7-tuple ($f_1\ f_2\ f_3\ f_4\ f_5\ f_6\ f_7$), providing detailed features. The augmented features are as follows;

f_4: grammatical class (shown above)
f_5: syntactical, original internal-structure
f_6: morphosyntactical variation: $f_6=(m_1, m_2, \ldots, m_9)$
- m_1: possibility to be modified by adnominal
- m_2: possibility to be modified by appredicative
- m_3: auxiliaries insertable in between its constituent words
- m_4: particles insertable in between its constituent words
- m_5: deletable particles
- m_6: particles by which those in it are replaced
- m_7: constituents which can be reordered
- m_8: possibility to be nominalized by inversion
- m_9: possibility to be passivized

f_7: estimated relative frequency

f_6 was adopted to ensure the flexibility of MWEs, while controlling the number of headings. Thus, our lexicon is not simply a rigid list of MWEs but designed as a resource proliferous to a total variety of idiosyncraticity. (Shudo et al., 1980, 1988; Shudo, 1989; Yasutake et al., 1997).

Although its development has not been completed, some preliminary experiments to apply this resource to NLP tasks have been reported in (Koyama et al., 1998; Iwase et al., 2000).

The present study focuses on a relatively small, but crucial subset, i.e., the set of NPCIs, including ordinary auxiliaries and ending-particles, and its relationship to the non-propositional structure of natural Japanese sentences[3].

3 Non-propositional Structures (NPSs)

Let us consider the meaning of a sentence;

(1) “彼・は・そこ・に・居る・べきで・なかっ・た *kare・ha・soko・ni・iru・bekide・nakat・ta*” ‘*He should not have been there*’,

where a verb “居る *iru*” ‘*be*’ is followed successively by three conventional auxiliaries, “べきだ *bekida*” ‘should’, “ない *nai*” ‘*not*’ and “た *ta*” ‘*-ed*’ which mean obligation, negation and past-tense, respectively, in the sentence-final position[4]. According to the sequence of their occurrences, the solely literal paraphrase of (1) would be something like;

(2) “彼・は・そこ・に・居る・べきだ・と・いう・こと・は・なかっ・た

kare・ha・soko・ni・iru・bekida・to・iu・koto・ha・nakat・ta” ‘*It was not necessary for him to be there*’,

However, this reading is not correct for (1). Rather, its regular reading should be something like;

(3) “彼・が・そこ・に・居・た・の・は・まずい
kare・ga・soko・ni・i・ta・no・ha・mazui” ‘*It is evaluated in the negative that he was there*’,

By the way, it will be reasonable to think sentences (2) and (3) share a kernel sentence “彼・が・そこ・に・居る *kare・ga・soko・ni・iru*” ‘*He is there*’, into which NPCs are incorporated successively, i.e., first-obligation, second-negation, third-past-tense, in the case of (2), and first - past-tense, second - speaker's-negative-evaluation, in the case of (3). Moreover, each stage of this incorporation would be regarded as mapping the utterance's meaning from one to another, in parallel with a syntactic form being mapped from one to another. Hence, by introducing Non-propositional Primitive Functions (NPFs), e.g., $\mathrm{OBLIGATION}_2$, $\mathrm{NEGATION}_1$, PAST-TENSE, and NEG-EVAL, we can explain the Non-propositional Structure (NPS) of (2) as;

(4) PAST-TENSE［$\mathrm{NEGATION}_1$［$\mathrm{OBLIGATION}_2$［“彼・が・そこ・に・居る *kare・ga・soko・ni・iru*” ‘*He is there*’]]]

and NPS of (3), hence, of (1) as,

(5) NEG-EVAL［PAST-TENSE［“彼・が・そこ・に・居る *kare・ga・soko・ni・iru*” ‘*He is there*’]].[5]

Here, a problem is that (4) is wrong for (1). In order to cope with this non-compositionality, while adopting a MWE, “べきで・なかっ・た *bekide・nakat・ta*” as a NPCI with a triplet (100) which has a composite NPF, NEG-EVAL［PAST-TENSE［x]][6], we have designed our segmenter to prefer a longer segment by the least-cost evaluation. It should be noted that a composite of NPFs like this could be associated with a single NPCI.[7] This is caused by its idiomaticity, i.e., by the difficulty in decomposing it into semantically consistent sub-forms.

Investigation of a reasonably sized set of Japanese linguistic data revealed that NPS of a natural Japanese sentence could be generally formulated as a nested functional form;

(6) $\mathrm{M_n}[\mathrm{M_{n-1}}...[\mathrm{M_2}[\mathrm{M_1}[\mathrm{S}]]]...]$,

where S is a propositional, kernel sentence; $\mathrm{M_i}$ ($1 \leq i \leq n$), a NPF.

In the following, we sometimes employ the notation for a composite function, $\mathrm{M_n} \circ \mathrm{M_{n-1}} ... \circ \mathrm{M_2} \circ \mathrm{M_1}$, where $\mathrm{M_n} \circ \mathrm{M_{n-1}} ... \circ \mathrm{M_2} \circ \mathrm{M_1}[\mathrm{S}] =_{\mathrm{def}} \mathrm{M_n}[\mathrm{M_{n-1}}...[\mathrm{M_2}[\mathrm{M_1}[\mathrm{S}]]]...]$

4 Non-propositional Content Indicators (NPCIs) and Non-propositional Functions (NPFs)

We have settled a set of 150 basic NPFs by classifying 1,500 NPCIs which had been extracted from the approximately 200,000 sentence data in unrestricted domains. During the extraction of NPCIs and settlement of NPFs, they have been continuously checked by comparing with various dictionaries and linguistic literature such as (Morita et al., 1989).

They are subclassified as follows. Each polysemous NPCI is shared by different subclasses.

Examples of NPCIs and the number of NPFs are given in brackets in the following.

$\mathrm{F_1}$: polarity ＜3＞:
$\mathrm{NEGATION}_1$(“ない *nai*” ‘*not*’; “の・で・は・ない (100) *no・de・ha・nai*” ‘*not*’; etc.), $\mathrm{NEGATION}_2$(“と・いう・訳・で・は・ない(100) *to・iu・wake・de・ha・nai*” ‘*not*’; etc.), etc.

$\mathrm{F_2}$: tense ＜1＞:
PAST-TENSE (“た *ta*” V-ed ; “だ *da*” V-ed)

$\mathrm{F_3}$: aspect-observational ＜9＞:
IMMEDI-AFT-TERMINATING (“た・ところ・だ(110) *ta・tokoro・da*” ‘*have just* V-en’; “た・ばかり・の・ところ・だ(110) *ta・bakari・no・tokoro・da*” ‘*have just* V-en’; etc.), IMMEDI-BEF-BEGINING(“う・と・し・て・いる(100) *u・to・si・te・iru*” ‘*be about to*’; “よう・と・し・て・いる(100) *you・to・si・te・iru*” ‘*be about to*’; etc.), $\mathrm{PROGRESSING}_1$ (“て・いる(100) *te・iru*” ‘be V-ing’; “つつ・ある(110) *tutu・aru*” ‘be V-ing’; etc.), etc.

$\mathrm{F_4}$: aspect-action ＜8＞:
INCHOATIVE (“はじめる *hajimeru*” ‘*begin to*’; “だす *dasu*” ‘*begin to*’; etc.), TERMINATIVE (“おわる

owaru" '*finish* V-ing'; "おえる *oeru*" '*finish* V-ing'; etc.), CONTINUATIVE ("続ける *tuzukeru*" '*continue to*'; "永らえる *nagaraeru*" '*continue to*'; etc), etc.

F_5 : voice <10>:
PASSIVE ("れる *reru*" 'be V-en'; "られる *rareru*" 'be V-en'), CAUSATIVE ("せる *seru*" '*make*...V...'; "させる *saseru*" '*make*...V...'), PASSIVE-SUFFERING ("れる *reru*" '*have*...V-en'; "られる *rareru*" '*have*...V-en'; etc.), PASSIVE-BENEFIT-TAKING$_1$ ("て・もらう (100) *te・morau*" '*ask*...V'; "て・いただく (100) *te・itadaku*" '*ask*... V'; etc.), BENEFIT-TAKING ("て・くれる (100) *te・kureru*" '(*someone*) V... *for* (*me*)...'; etc.), BENEFIT-GIVING ("て・やる (100) *te・yaru*" '(*I*) V... *for* (*someone*)...'; etc.), 'etc.

F_6 : politeness-operator <3>:
POLITENESS$_1$ ("ます *masu*"; etc.), etc.

F_7 : predicate-suffix <30>:
TRIAL ("て・みる (100) *te・miru*" '*try to*'; etc.), etc.

F_8 : modality <60>:
NEG-EVAL ("べき・で・ない (100) *beki・de・nai*" '*should not*'; "の・は・よく・ない (100) *no・ha・yoku・nai*" '*should not*'; etc.), OBLIGATION$_2$ ("必要・がある (100) *hituyou・ga・aru*" '*need*', "べきだ *bekida*" '*should*', etc.), OBLIGATION$_1$ ("なければ・ば・なら・ない (111) *nakere・ba・nara・nai*" '*have to*'; etc.), PROHIBITION ("て・は・なら・ない (111) *te・ha・nara・nai*" '*should not*', etc.), CAPABILITY ("得る *uru*"; "こと・が・できる (100) *koto・ga・dekiru*" '*be able to*'; etc.), GUESS$_1$ ("う *u*" '*will*'), GUESS$_3$ ("かも・しれ・ない *kamo・sire・nai*" '*may*'), etc.

F_9 : illocutionary-act <28>:
IMPERATIVE (verb's 'imperative form'), INTERROGATIVE ("か *ka*" '*interrogative form*'; "の・か (110) *no・ka*" '*interrogative form*'; etc.), PROHIBITIVE ("な *na*" '*Don't*...'), PERMISSIVE ("て・よい (100) *te・yoi*" 'You may...'; "て・も・かまわ・ない (100) *te・mo・kamawa・nai*" 'You may...'; etc.), REQUESTING ("て・くれ (110) *te・kure*" '*Please*...'; "て・ほしい (110) *te・hosii*" '*I want you to*...'; etc.), etc.

MWEs are essential for indicating the NPC of Japanese sentences. For instance, without MWEs, "なければ・ば・なら・ない (111) *nakere・ba・nara・nai*" and its family in F_8 which are equivalent to English '*have to*', being used quite frequently in texts, we could not indicate OBLIGATION1 i.e., 'strong obligation' in a simple Japanese sentence. Instead, we would be forced to construct a somewhat verbose structure such as "...こと・が・強く・要求・される *...koto・ga・tuyoku・youkyuu・sareru*" '*It is strongly required that*...', without them.

5 Identification of NPSs

5.1 Sentence-final Structure in Japanese

Employing MWEs as NPCIs enabled us to describe the outermost, basic structure of a Japanese sentence by the following production rules;

(7) $S_0 \rightarrow BP^* \cdot PRED$,
(8) $S_i \rightarrow S_{i-1} \cdot m_i$, $(1 \leq i \leq n)$,

where S_0 denotes a kernel sentence; BP, a basic phrase called *bunsetsu*; PRED, a predicate of the kernel sentence; S_i, a sentence, m_i, a NPCI and a symbol '*', closure operator on the concatenation, '·'. In the following, we sometimes use predicative parts, $PRED \cdot m_1 \cdot m_2 \cdot \ldots \cdot m_n$ instead of full sentences, for simplicity.

5.2 Occurrence of MWEs as NPCIs

Here we show a statistical data in Table 1 to clarify how often NPCIs and multiword NPCIs are used in the predicative part, $PRED \cdot m_1 \cdot m_2 \cdot \ldots \cdot m_n$, of actual Japanese sentences. The 9,210 test predicative parts whose PREDs are verbal or adjectival are brought randomly from EDR-corpus (EDR, 1996), which are independent of the extraction of NPCIs. Table 1 gives the total number of single-word NPCIs and multiword NPCIs occurring in the predicative parts for each n. The overall rate for MWE usage is approximately 0.42. The rate for the at-least-once-occurrence of NPCI is 0.47. The maximum value of n was 5.

Table 1 Statistics on NPCIs and MWEs in predicative part; $PRED \cdot m_1 \cdot m_2 \cdot \ldots \cdot m_n$

n	occurrences of predicative parts	occurrences of non-MWEs (A)	occurrences of MWEs (B)	A+B	B/(A+B)
0	4,899	—	—	—	—
1	3,131	1,852	1,279	3,131	0.408
2	966	1,128	804	1,932	0.416
3	178	276	258	534	0.483
4	34	63	73	136	0.537
5	2	7	3	10	0.300
total	9,210	3,326	2,417	5,743	0.421

Here, verbal PREDs are divided into two types; verb-type and verbal-noun・“する *suru*”-type. Adjectival PREDs are also divided into two types; adjective-type and adjectival-noun・“だ *da*”-type.

5.3 NPS Analysis-Experiment and Result

Our morphology model adopts MWEs as atomic elements for sentence construction and the bi-gram at syntactic-category-level. It is formally a probabilistic finite automaton with 150 states that prescribes minutely the internal structure of each BP and the predicative part, PRED・m_1・m_2・…・m_n.

On the basis of this model, we have developed a pilot system consisting of a segmenter (SEG) that segments the input predicative part into a PRED and each NPCI using the least-cost method, and a NPS-constructor (NPSC) that constructs NPSs of the segmented predicative part. The system is not for actual working in society, at present, but for certifying the fundamental validity of our NPF framework and for studying for its improvement.

SEG outputs all least-cost solutions for a given input. Also, NPSC produces all possible NPSs for a predicative part which includes polysemous NPCI's.

For example, SEG segments an input,

(9) “読まなければならないだろう
yomanakerebanaranaidarou” ‘*will have to read*’
into
(10) “読ま/・なけれ・ば・なら・ない(111)/・だろ・う(011)
yoma/・nakere・ba・nara・nai/・daro・u”,

where a slash ‘/’ denotes a segment-boundary identified by SEG.
NPSC evaluates a function *nps* defined below.

(11) $nps(S_0)=S_0$,
$nps((S_0/m_1/m_2.../m_i)=M_i^k[...M_i^2[M_i^1[nps(S_0/m_1/m_2.../m_{i-1})]]]$, $(1\leq i\leq n)$,

where $M_i^k[...M_i^2[M_i^1[x]]]$ is a NPF (if $k=1$) or a composite of NPFs (if $k\geq 2$) associated with m_i in NPCI lexicon.
For example, the computation of *nps* for (10) is;

(12) *nps*(“読ま/・なけれ・ば・なら・ない(111)/・だろ・う(011)*yoma/・nakere・ba・nara・nai/・daro・u*”)
$=GUESS_2$[*nps*(“読ま/・なけれ・ば・なら・ない(111) *yoma/・nakere・ba・nara・nai*” ‘*have to read*’)]
$=GUESS_2$[$OBLIGATION_1$[*nps* (“読む *yomu*” ‘*read*’)]]
$=GUESS_2$[$OBLIGATION_1$[“読む *yomu*” ‘*read*’]],

where NPFs, $GUESS_2$ and $OBLIGATION_1$ are associated with “だろ・う *daro・u*(011)” ‘*will*’ and “なけれ・ば・なら・ない *nakere・ba・nara・nai* (111)” ‘*have to*’, respectively.

We have evaluated 4,418 outputs of SEG, for 4,413 input predicative parts taken randomly from EDR-corpus in unrestricted domains (i.e., originally news paper articles). The input set is independent of that used in the development of NPF framework and that used in the examination described in Section 5.2. 4,083 outputs are evaluated as correct. Thus, SEG produced a recall of 0.93 and a precision, 0.92. SEG produced an unique output for most of the inputs.

Most of unacceptable outputs of SEG are caused by the ‘over-identification’ of MWEs, namely, they contain MWEs which should be ‘sub-segmented’ for those inputs. For example, a multiword NPCI, “た・もの・だ (110) *ta・mono・da*” ‘*used to*’ indicating CUSTOM-PAST had been settled for the case that the original meaning of a noun, “もの *mono*” ‘*thing*’ is not completely ‘alive’ like in “私・は・よく・そこで・花・を・買っ・た・もの・だ *watasi・ha・yoku・soko・de・hana・wo・kat・ta・mono・da*” ‘*I often used to buy a flower there*’. However, it is alive in a sentence “この・花・は・君・が・そこ・で・買っ・た・もの・だ *kono・hana・ha・kimi・ga・soko・de・kat・ta・mono・da*” ‘*This flower is what* (*that which*) *you bought there*’. For this sentence, SEG is not able to output the correct segmentation, i.e., “た/・もの/・だ *ta/・mono/・da*”, since the cost is heuristically given to the system so that it prefers longer segments.

It will be necessary in future to contrive some mechanism to cope with this problem, e.g., by selecting one correct solution after trying both, if necessary, MWE-based case and sub-segmented case. This is an essential and inevitable problem which arises in all NLPs which adopt a large-scale set of MWEs.

Another cause, which is miner, for the errors is the ‘missing’ of the multiword NPCI in the lexicon. For example, we found that “た・の・だっ・た *ta・no・da*

t・ta", which was missing in our lexicon, occurred in "彼・は・聞い/・た/・の・だっ/・た *kare・ha・kii/・ta/・no・da/t・ta*" '*He asked.*' should be a multiword NPCI marked with (110), for DECLARATIVE ○ PAST-TENSE, since the duplication of PAST-TENSE marker, "た *ta*" in this case should not be treated as it is, but it should be treated as a single negation in the processing.

Next, for 4,083 predicative parts correctly segmented by SEG, we examined the outputs of NPSC. It turned out that the total number of outputs and the number of correct outputs are 9,502 and 3,975, respectively. That is, NPSC produced a recall of 0.97 and a precision of 0.42. Overall recall, i.e., to be produced by NPSC combined with SEG would be 0.90. While the score of the recall roughly suggests the sufficiency of the collection of NPFs, we found some candidate NPFs missing in our list. For instance, although OBLIGATION$_3$ in $\mathbf{F}_8$ is associated with "もの・だ(111) *mono・da*" in the lexicon for the case like "法・に・は・従う/・もの・だ *hou・niha・sitagau/・mono・da*" '(*You*) *ought to obey the law*', GENERAL-IDEA for "父親・は・優しい/・もの・だ *chichioya・ha・yasasii/・mono・da*" '*In general, a father is gentle*' and IMPRESSED for "彼・は・よく・働く/・もの・だ *kare・ha・yoku・hataraku/・mono・da*" '*How hard he works!*' were missing in our list. This kind of the absence of NPFs in the system is the dominant cause of the errors NPSC produced.

Relatively low score of the precision is due to the system's 'over-generation' mainly caused by the semantic ambiguities of NPCIs. For example, "て・いる(100) *te・iru*", which is frequently used in daily documents or speech, is semantically ambiguous. Three examples of NPSs which NPSC produces for an input, "彼・が・食べ/・て・いる (100) *kare・ga・tabe/・te・iru*" are;

(13) *nps*("彼・が・食べ/・て・いる(100) *kare・ga・tabe/・te・iru*")
=$_1$ PROGRESSING$_1$["彼・が・食べる *kare・ga・taberu*" '*He eats*'],
=$_2$ PROGRESSING$_2$["彼・が・食べる *kare・ga・taberu*" '*He eats*'],
=$_3$ COMPLETED$_1$["彼・が・食べる *kare・ga・taberu*" '*He eats*'],

where each output corresponds to, '*He is eating*', '*He has been eating*' and 'He has eaten' , respectively. It will be impossible, in this case, to disambiguate these meanings of "て・いる(100) *te・iru*" by any process within the input sentence. We make a thing of the recall rather than the precision for baseline systems composing a large NLP system, since we believe that narrowing the possibilities toward the best solution is easier than broadening the possibilities so as to include the correct one. We could go so far as to say that NLP should be, in a long run, multi-layered process of 'disambiguation'. Needless to say, in the earlier layers in NLP we accomplish the pruning, the more effectively we can suppress their increase which might be explosive in subsequent, higher layers. For example, for an input "彼・が・食べ/・始め/・て・いる(100) *kare・ga・tabe/・hajime/・te・iru*" '*He has begun to eat*', NPSC also produces three outputs, which correspond to '*He is beginning to eat*', '*He has been beginning to eat*' and '*He has begun to eat*', respectively, shown in (14), according to the ambiguity of "て・いる(100) *te・iru*" mentioned above.

(14) *nps*("彼・が・食べ/・始め/・て・いる(100) *tabe/・hajime/・te・iru*")
=$_1$ PROGRESSING$_1$[INCHOATIVE ["彼・が・食べる *kare・ga・taberu*" '*He eats*']],
=$_2$ PROGRESSING$_2$[INCHOATIVE ["彼・が・食べる *kare・ga・taberu*" '*He eats*']],
=$_3$ COMPLETED$_1$[INCHOATIVE ["彼・が・食べる *kare・ga・taberu*" '*He eats*']],

However, a NPCI "始める *hajimeru*" for INCHOATIVE, corresponding to English 'begin to' is used in this case. This might enable the system to exclude the first two inappropriate interpretations by aspect-feature constraints which prohibit functional compositions, PROGRESSING$_1$○ INCHOATIVE and PROGRESSING$_2$ ○ INCHOATIVE. In fact, a predicate having 'instantaneous' aspect can not become 'progressive'. Thus, semantic constraints or preferences to control the composition operation '○' of NPFs will be effective for disambiguating the meanings of NPCI within our framework, and hence, to produce a better precision. However, the detailed measures of this type as well as more sophisticated, powerful measures for the disambiguation based on

more global information, are left to future work.

Generatively, most NPFs simply command to put their proper NPCI in the final position of the given sentence in Japanese. However, some NPFs, such as CAUSATIVE, PASSIVE require additional restructuring of the given sentence on the basis of additional information about such as agent, recipient, etc. Then, analytical process for these NPFs should include rewriting of the given sentence back to the original in order to decide the correct kernel sentence. NPSC itself is not designed to accomplish this process but designed to decide just a sequence of NPFs. We describe briefly this process in Appendix.

6 On Some Applications

6.1 Machine Translation

It will be interesting whether our NPS framework works as a 'pivot' for machine translation (MT) of non-inferencial, non-propositional sentences. That is whether it can be language-free or not. As shown in Section 5.3, as far as the analytical processing of Japanese sentence is concerned, the framework seems to be fundamentally feasible, though there left a problem of disambiguation. On the other hand, with respect to the generative phase for target languages other than Japanese, we can not say definitely anything without deep linguistic knowledge of those languages. We just introduce briefly here a small, pilot experimental system, referred to as ENGL, whose input is the NPS of a sentence and whose output is its English surface forms. ENGL simply realizes NPFs within English syntax. We assumed temporarily each NPF for English could be accomplished by applying rewriting rules of two types; i) V→x･V_v･y and ii) S →x･S_v･y , where V is a verb or an auxiliary; V_v is V, a null string, or a variant of V; S, a sentence; S_v, a variant of S; and x, y, a null string or a string of specific words. Several rules, if necessary, are associated with a single NPF to be applied selectively or successively. In addition, no rule is associated with a NPF in some cases. For example, POLITENESS, which is common in colloquial Japanese, has mostly no NPCI in English, so is associated with a 'null' operation. ENGL tries to produce all possible English predicative forms applying all applicable rewriting rules associated with a given NPF.

We will show some examples of the computation executed by ENGL. The underlined word in the following is the main verb (or auxiliary) to which the next rewriting rule can be applied.

(12') *nps*(“読ま/･なけれ･ば･なら･ない(111)/･だろ･う(011) *yoma/･nakere･ba･nara･nai/･daro･u*”)
＝GUESS$_2$ [OBLIGATION$_1$ [“読む *yomu*”]]
＝GUESS$_2$ [OBLIGATION$_1$ [*read*]]
＝GUESS$_2$ [*have to* ･ *read*]
＝*will･have to･read*,

where the rewriting rules associated with OBLIGATION$_1$ and GUESS$_2$ are V→*have to* ･ V_{root} and V→*will* ･ V_{root}, respectively.

(15) *nps*(“読ん/･で･も･よい(110)/･かも･しれ･ませ･ん(111)/･ね *yon/･de･ mo･yoi/･kamo･sire･mase･n/･ne*”)
＝MAKING-SURE [POLITENESS [GUESS$_3$ [PERMISSIVE [“読む *yomu*”]]]]
＝MAKING-SURE [POLITENESS [GUESS$_3$ [PERMISSIVE [‘*read*’]]]]
＝MAKING-SURE [POLITENESS [GUESS$_3$ [‘*It is allowed that ... read*’]]]
＝MAKING-SURE [POLITENESS [*It may be allowed that...read*]]]
＝MAKING-SURE [*It may be allowed that...read*]
＝*It may be allowed that...read..., right?*

where the rules for PERMISSIVE, GUESS$_3$, POLITENESS and MAKING-SURE are, S→*It is allowed that* ･ S, V → *may* ･ V_{root}, S → S and S → S ･, *right?*, respectively.

ENGL is incorporated with an aspectual gadget, similar to that introduced in the discussion on the example (14), to disambiguate the meaning of NPCI, “て･いる(100) *te･iru*”. By this, the 'static' aspect of a verb, “愛する *aisuru*” '*love*' excludes the possibilities other than STATE-OF-THINGS in (16).

(16) *nps*(“愛し/･て･いる(100) *aisi/･te･iru*”)
＝STATE-OF-THINGS [*love*]＝love

In a small experiment, for 300 NPSs extracted from sentences in technical papers, which are independent of the rule development, ENGL produced a precision of 0.86 and a recall, 0.80. These relatively high scores are caused by that we excluded CAUSATIVE and PASSIVE, which require additional information to be realized in surface forms in the evaluation and that technical papers tend to have less-complicated NPSs. While it suggests the fundamental validity of the NPF framework, more extensive tests will be required in future to make more reliable evaluation for the general domain. In addition, further correction and refinement of rewriting rules and their applicability conditions for English will be essential to obtain higher scores.

6.2 Formal Description of Similarity between NPSs

In this section, we show that our framework for the NPS description could be used properly to formalize some semantic or pragmatic relationship between non-propositionalized sentences.

6.2.1 Logical Rules

First, we discuss, here, the logical similarity relation, $\rightleftharpoons \subset ((\cup \mathbf{F}_i)^*)^2$, $(1 \leq i \leq 8)$, between NPSs which seems crutial for NLP tasks such as text retrieval or paraphrasing.[8] We prefer the term, 'similarity' to 'equivalence', here, since the relationship should be graded for such NLP tasks as text retrieval. More theoretically, it should be based on truth values taken in 'most situations', or in some 'similar' worlds by the sentences.[9]

In the course of investigation of NPCIs and NPFs, it turned out that there are several basic rules such as;

(17) $\text{NEG-EVAL} \circ \text{NEGATION}_1 \rightleftharpoons \text{OBLIGATION}_2$,
(18) $\text{NEGATION}_1 \circ \text{PERMISSION} \rightleftharpoons \text{PROHIBITION}$,
(19) $\text{NEGATION}_{1,2} \circ \text{NEGATION}_{1,2} \rightleftharpoons \lambda$ (identity function),
(20) $N \circ \lambda \rightleftharpoons \lambda \circ N$ for $^{\forall}N \in (\cup \mathbf{F}_i)^*$, $(1 \leq i \leq 8)$,
(21) $\text{POLITENESS} \rightleftharpoons \lambda$.

(17) asserts that, for example, an utterance, "*He has to go there.*" is similar to "*It is evaluated in the negative that he does not go there.*". Besides these basic rules, there is a notable set of logically rules. For example, from the observation that "働い/・て・ばかり・いる(110)/・訳・で・は・ない(110) *hatarai/・te・bakari・iru/・wake・de・ha・nai*" '*do not always work*' is similar to "働か・ない/・時・が・ある(010) *hataraka・nai/・toki・ga・aru*" '*It happens occasionally that...do not work*' the following rule will be induced;

(22) $\text{NEGATION}_2 \circ \text{HIGHEST-FREQUENCY} \rightleftharpoons \text{LOW-FREQUENCY} \circ \text{NEGATION}_1$.

Also, "働か/・なく/・て・も・よい(110) *hataraka/・naku/・te・mo・yoi*" '*need not work*'; '*It is allowed that...do not work...*' and its paraphrase, "働か/・なけれ・ば・なら・ない (111) /・事・は・ない (100) *hataraka/・nakere・ba・nara・nai/・koto・ha・nai*" '*It is not obligatory that... work...*', might induce a rule;

(23) $\text{PERMISSION} \circ \text{NEGATION}_1 \rightleftharpoons \text{NEGATION}_1 \circ \text{OBLIGATION}$.

These rules can be generalized as (24), (24') by introducing a 'duality' function, $\boldsymbol{d}$ defined in Table 2,

Table 2 Definition of 'duality' function $\boldsymbol{d}$

$\boldsymbol{M}$, $\boldsymbol{d}(\boldsymbol{N})$	$\boldsymbol{d}(\boldsymbol{M})$, $\boldsymbol{N}$
POSSIBILITY	NECESSITY, HIGHEST-PROBABILITY, HIGHEST-CERTAINTY
LOW-FREQUENCY	HIGHEST-FREQUENCY, HIGHEST-USUALITY
PERMISSION	OBLIGATION, HIGHEST-INEVITABILITY

(24) $\text{NEGATION}_{1,2} \circ \boldsymbol{M} \rightleftharpoons \boldsymbol{d}(\boldsymbol{M}) \circ \text{NEGATION}_{1,2}$,
(24') $\boldsymbol{M} \rightleftharpoons \text{NEGATION}_{1,2}\, \boldsymbol{d}(\boldsymbol{M}) \circ \text{NEGATION}_{1,2}$.

We show two more examples;

(22') $\text{HIGHEST-FREQUENCY} \circ \text{NEGATION}_1 \rightleftharpoons \text{NEGATION}_2 \circ \text{LOW-FREQUENCY}$.
nps("働か/・ない・で・ばかり・いる(111) *hataraka/・nai・de・bakari・iru*" '*It is always that...do not work...*') $\rightleftharpoons$ *nps*("働く/・こと・が・ある(100)/・と・は・言え・ない(110) *hataraku/・koto・ga・aru/・to・ha・ie・nai*" '*It does not happen that...sometimes work...*').

(23') $\text{OBLIGATION} \circ \text{NEGATION}_1 \rightleftharpoons \text{NEGATION}_1 \circ \text{PERMISSION}$.

nps(“働い/･て･は･なら･ない(111) *hatarai/･te･ha･nara･nai*” ‘*must not work*’) ≒ *nps*(“働い/･て･よい(100)/･と･いう･事･は･ない(100) *hatarai/･te･yoi/･to･iu･koto･ha･nai*” ‘*It is not permissible that...work...*’).

Rule (24) corresponds to the axiom, $\neg\Diamond P = \Box\neg P$, in modal logic and its variants, e.g., deontic (Mally, 1926) or temporal (Prior, 1967) logic, where $\Box$ and $\Diamond$ are the necessity and possibility operator, respectively.

6.2.2 Pragmatic Rules

The similarity relation among the speaker's attitude or intention toward the hearer is defined as a set, $\backsim\subset\{(a,b)\,|\,a,b\in(\mathbf{F}_1\cup\mathbf{F}_2\ldots\cup\mathbf{F}_9)^*\wedge((^{\exists}i,1\leq i\leq 1\wedge f_i\in\mathbf{F}_9)\vee(^{\exists}j,1\leq j\leq m\wedge g_i\in\mathbf{F}_9))$, where $a=f_1\circ f_2\ldots\circ f_l$, $b=g_1\circ g_2\ldots\circ g_m\}$.

Some of the *indirect speech acts* (Searle, 1975) can be formulated as the similarity within our framework. Examples of the rules and their instances are;

(25) REQUESTING
- $\backsim$INTERROGATIVE$\circ$NEGATION$_1\circ$ CAPABILITY,
- $\backsim$INTERROGATIVE$\circ$CAPABILITY,
- $\backsim$INTERROGATIVE$\circ$NEGATION$_1\circ$BENEFIT-TAKING,
- $\backsim$INTERROGATIVE$\circ$NEGATION$_1\circ$CAPABILITY$\circ$PASSIVE-BENEFIT-TAKING,
- $\backsim$DESIRING$\circ$PASSIVE-BENEFIT-TAKING,
- $\backsim$DESIRING$\circ$PASSIVE.

nps(“見/･て･くれ(110) *mi/･te･kure*” ‘*Look at ...*’),
$\backsim$*nps*(“見る/･こと･が･出来 (100) /･ない/･か *miru/･koto･ga･deki/･nai/･ka*” ‘*Can't you look at ...?*’),
$\backsim$*nps*(“見る/･こと･が･出来る (100) /･か *miru/･koto･ga･dekiru/･ka*” ‘*Can you look at ...?*’),
$\backsim$*nps*(“見/･て･くれ(110)/･ない/･か *mi/･te･kure/･nai/･ka*” ‘*Don't you look at ... for me ...?*’),
$\backsim$*nps*(“見/･て･もら(100)/･え/･ない/･か *mi/･te･mora/･e/･nai/･ka*” ‘*Can't I have you look at... for me...?*’),
$\backsim$*nps*(“見/･て･もらい(100)/･たい *mi/･te･morai/･tai*” ‘*I want you to look at ... for me...*’),
$\backsim$*nps*(“見/･られ/･たい *mi/･rare/･tai*” ‘*I want you to look at ...*’).

(25) asserts that, for example, an *indirect speech*, “これ･を･見る/･こと･が･出来(100)/･ない/･か *kore･wo･miru/･koto･ga･deki/･nai/･ka*” ‘*Can't you look at this?*’ can be occasionally used in place of “これ･を･見/･て･くれ(110)*kore･wo･mi/･te･kure*” ‘*Look at this*’. These utterances are similar in the sense that both transmit the speaker's expectation of the same hearer's responses but its strength, speaker's intention or concern involved in each are different.

With respect to prohibition, invitation, permission and assertion, we will have;

(26) PROHIBITIVE
- $\backsim$PROHIBITION,
- $\backsim$NEGATION$_1\circ$CAPABILITY.

nps(“入る/･な *hairu/･na*” ‘*Do not enter...*’)
- $\backsim$*nps* (“入っ/･て･は･なら･ない (111) *hait/･te･ha･nara･nai*” ‘*You must not enter...*’),
- $\backsim$*nps*(“入る/･事･が･出来(101) /･ない *hairu/･koto･ga･deki/･nai*” ‘*You can not enter...*’),

(27) INVITING
- $\backsim$INTERROGATIVE$\circ$INVITING,
- $\backsim$INTERROGATIVE$\circ$NEGATIVE$_1$.

nps(“食べよ/･う *tabeyo/･u*” ‘*Let's eat...*’)
- $\backsim$*nps*(“食べよ/･う/･か *tabeyo/･u/･ka*” ‘*Will you eat...?*’),
- $\backsim$*nps*(“食べ/･ない/･か *tabe/･nai/･ka*” ‘*Don't you eat...?*’).

(28) PERMISSIVE
- $\backsim$POSSIBILITY.

nps(“着/･て･よい(100) *ki/･te･yoi*” ‘*You may wear...*’)
- $\backsim$*nps*(“着る/･こと･が･出来る(100) *kiru/･koto･ga･dekiru*” ‘*You can wear...*’).

(29) ASSERTING$\circ$PAST-TENSE$\circ$NEGATION$_1$
- $\backsim$INTERROGATIVE$\circ$PAST-TENSE

nps(“食べ/･なかっ/･た/･よ *tabe/･nakat/･ta/･yo*”; ‘*I did not eat...*’),
$\backsim$*nps*(“食べ/･た/･かい *tabe/･ta/･kai*”; ‘*Did I eat ...?*’).

Some NPFs such as POLITENESS, MODESTY and HONORIFICITY, lexically expressed often in Japanese, expand these rules. For example, a rule; REQUESTING $\backsim$ INTERROGATIVE$\circ$NEGATION$_1\circ$ BENEFIT-TAKING, in (25) will be expanded to a rule;
REQUESTING $\backsim$ INTERROGATIVE$\circ$NEGATION$_1\circ$ POLITENESS$\circ$MODESTY$\circ$BENEFIT-TAKING for

an utterance, "これを・見/・て・ください(110)/・ませ/・ん/・か *kore・wo・mi/・te・kudasai/・mase/・n/・ka*" '*Would you mind looking at this (for me)?*', expressing more modest and polite request.

At present, we have approximately 30 pragmatic rules which are basic in this sense, concerned with the NPCIs in Japanese. In the realistic tasks of NLP, application of these rules should be controlled by rather complicated conditions settled for each of them. For example, conditions for rules (25)～(28) will include that the agent of their complement sentence should be the second person, and for (29), the first. Although the principle underlying these rules were discussed in a lot of literature, e.g., *felicity condition* in (Searle, 1975), etc., the whole picture has not been clarified for computational usage.

7 Conclusion

We have shown that as far as the non-inferential, Non-Propositional Content (NPC) in Japanese sentence is concerned, its semantic compositionality can be secured, provided sentence-final MWEs are adopted properly as NPCIs. Although the functional treatment of NPCs is not particularly new in the theoretical domain, our model is characterized by its broad syntactic/semantic coverage and its tractability in NLP. It connects syntax with semantics by actually defining 150 non-propositional functions (NPFs) for 1500 NPC indicators through a large-scale empirical study. The similarity equations presented in Section 6 might lead to some formal system of 'calculations' on the set of NPFs, which might be available for NLP in future.

The syntactic coverage of our semantic/pragmatic model should be further broadened by investigating non-final parts of Japanese sentences. This research would focus on the sentence embedment whose main verb is epistemic or *performative* (Austin, 1962), and adverbs that take part in indicating NPCs.

While the list of NPFs introduced in this paper will provide, we believe, a basis for analyzing the NPC of natural sentences, it might be possible, or rather necessary for particular task, to refine NPFs by enriching them with case-elements, more detailed degrees or subtle differences in meaning, etc.

We have not solved the problem of semantically disambiguating each NPCI. Further, we know little about the language-dependency, consistency of the similarity rules. The language-dependency of NPS is interesting from the viewpoint of machine translation or comparative pragmatics. The framework presented here could hopefully provide a tool for those comparative studies.

Acknowledgement

The authors would like to thank the late Toshihiko Kurihara for first stimulating this research. We also thank all people involved in our investigation of big amount of Japanese linguistic data.

Appendix

Syntactically, most NPFs (or composites of NPFs) simply direct rewriting of the form (8) in Japanese. However, PASSIVE or CAUSATIVE exceptionally directs rewriting of case-particles, "が *ga*" (subjective), "を *wo*" (objective) and "に *ni*" (dative) in S.

We outline the analysis process for PASSIVE and CAUSATIVE, adopting an example which includes their composite.

Let the input sentence be s="A・が・B・に・C・を・食べ・させ・られる *A・ga・B・ni・C・wo・tabe・sase・rareru*" '*A is made by B to eat C*'. From NPCIs, "させ(る) *sase(ru)*" (causative), "られる *rareru*" (passive) in the sentence-final, predicative part, "食べ・させ・られる *tabe・sase・rareru*", its outer-structure of NPS is determined as PASSIVE [CAUSATIVE [X]] by NPSC. Here, the kernel X is computed as follows;

X=1/CAUSATIVE [1/PASSIVE [s]]
=1/CAUSATIVE ["B・が・A・に・C・を・食べ・させる *B・ga・A・ni・C・wo・tabe・saseru*" '*B makes A eat C*']
="A・が・C・を・食べる *A・ga・C・wo・taberu*" '*A eats C*'.

That is, first, the inverse function 1/PASSIVE directs to delete "られる *rareru*" from the sentence and to record that A marked with "が *ga*" is the passive item. We have several choices here. If B is chosen as the active item, and "B・に *ni*" and "A・が *ga*" are rewritten as "B・が *ga*" and "A・に *ni*", respectively, then "B・が・A・に・C・を・食べ・させる *B・*

ga・*A*・*ni*・*C*・*wo*・*tabe*・*saseru*” ‘*B makes A eat C*’ is obtained as an intermediate solution. Next, 1/CAUSATIVE directs to delete “させる *saseru*” and “B・が *ga*”, recording that B marked with “が *ga*” is the agent of cause-action. Then, “A・に *ni*” is rewritten as “A・が *ga*”, recording that A is the patient of the cause-action. Thus, having NPFs enriched with the cases, subject and dative, a resulting NPS will be,

PASSIVE$_{(A,B)}$[CAUSATIVE$_{(B,A)}$[“A・が・C・を・食べる *A*・*ga*・*C*・*wo*・*taberu*” ‘*A eats C*’]].

Less feasible but possible interpretations for this input, such as; PASSIVE$_{(A,x)}$[CAUSATIVE$_{(A,B)}$ [“B・が・C・を・食べる *B*・*ga*・*C*・*wo*・*taberu*” ‘*B eats C*’]] corresponding to ‘A *is made* (*by someone*) *to cause* B *to eat* C’, or PASSIVE$_{(A,x)}$ [CAUSATIVE$_{(x,B)}$ [“B・が・C・を・食べる *B*・*ga*・*C*・*wo*・*taberu*” ‘*B eats C*’]] corresponding to ‘A *is suffered from that* (*someone*) *causes* B *to eat* C’, will be obtained in this processing as well. The stimulous problem of reducing the ambiguity of this type seems to be still open. For an original version of this procedure, see (Shudo et al., 1979).

Foonotes

(1) At present, f_1 will not be decided by any statistical method.

(2) An idea to treat compound nouns in Japanese language processing is presented in (Miyazaki et al., 1993).

(3) Some of multiword NPCIs are treated in a framework for MT in (Shirai et al., 1993).

(4) “べきだ *bekida*” and “ない *nai*” are inflected as “べきで *bekide*” and “なかっ *nakat*”, respectively, in (1).

(5) We use lower-suffixes to distinguish NPFs by subtle differences in meaning, degree, etc.

(6) Another choice could be, first, to adopt a shorter MWE, “べきで・ない *bekide*・*nai*” ‘*should not*’ as a NPCI indicating PROHIBITION$_2$, second, to build a NPS, PAST-TENSE[PROHIBITION$_2$[“彼・が・そこ・に・居る *kare*・*ga*・*soko*・*ni*・*iru*” ‘*He is there*’]], and last, to apply the following similarity rule in order to obtain (5), unless it yields the overgeneralization; PAST-TENSE[PROHIBITION$_2$ [x]] ⇌ NEG-EVAL[PAST-TENSE[x]]. The similarity rules are discussed in Section 6.

(7) Another typical example is “まい *mai*” which is a single auxiliary but has the meaning of ‘*will not*’, i.e., GUESS$_2$[NEGATION$_1$[x]].

(8) While the NPF in $\mathbf{F}_i$, ($1 \leq i \leq 7$) produces a truth conditional sentence, the NPF in $\mathbf{F}_9$ does not. The NPF in $\mathbf{F}_8$ produces a truth conditional sentence unless it is used for the speaker’s epistemic judgment.

(9) We do not enter further theoretical arguments here.

References

John L. Austin. 1962. How to Do Things with Words. Oxford U.P.

Osamu Iwase, Takumi Morimoto, and Kosho Shudo. 2000. Incorporating Collocation Data into a Statistical Language Model. Technical Report of IEICE: SP2000-113: 109-114. (in Japanese)

H. Paul Grice. 1975. Logic and Conversation. In P. Cole and J. L. Morgan, editors, Syntax and Semantics Vol. 3, Speech Acts: 41-57. Academic Press.

EDR (Japan Electronic Dictionary Research Institute). 1996. EDR Electronic Dictionary. http://www.iijnet.or.jp/edr/

Yasuo Koyama, Masako Yasutake, Kenji Yoshimura and Kosho Shudo.1998. Large-Scale Collocation Data and Their Application to Japanese Word Processor Technology. The Proc. of the 17th COLING: 694-698.

Ernst Mally. 1926. Grundgesetze des Sollens: Elemente der Logik des Willens. Universitats-Buchhandlung: viii＋85.

Karl Menger. 1939. A logic of the doubtful: On optative and imperative logic , in Reports of a Mathematical Colloquium: 2nd series: 2nd issue: 53-64. Notre Dame, Indiana University Press.

Masahiro Miyazaki, Satoru Ikehara and Akio Yokoo. 1993. Combined Word Retrieval for Bilingual Dictionary Based on the Analysis of Compound Words. Trans. IPSJ 34-4: 743-754. (in Japanese)

Yoshiyuki Morita and Masae Matsuki. 1989. Expression Pattern of Japanese. ALC Press. (in Japanese)

Arthur Prior. 1967. Past, Present and Future. Clarendon press, Oxford.

Iwan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. The Proc. of the 3rd CICLING: 1-15.

John R. Searle. 1975. Indirect Speech Acts. In P. Cole and J. L. Morgan, editors, Syntax and Semantics Vol. 3, Speech Acts: 59-82, Academic Press.

Satoshi Shirai, Satoru Ikehara and Tsukasa Kawaoka. 1993. Effects of Automatic Rewriting of Source Language within a Japanese to English MT System. Fifth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-93: 226-239]

Kosho Shudo, Toshiko Narahara and Sho Yoshida. 1979. A Study on the Canonical Form of Japanese Sentence. Joint Conference of Electrical and Electronics Engineers, Kyushu. (in Japanese)

Kosho Shudo, Toshiko Narahara and Sho Yoshida. 1980. Morphological Aspect of Japanese Language Processing. The Proc. of the 8th COLING: 1-8.

Kosho Shudo, Kenji Yoshimura, Mitsuno Takeuti and Kenzo Tsuda. 1988. On the Idiomatic Expressions in Japanese Language－An Approach through the Close and Extensive Investigation of Non-standard Usage of Words－IPSJ SIG Notes, NL-66-1: 1-7. (in Japanese)

Kosho Shudo. 1989. Fixed Collocations. Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research, 63101005. (in Japanese)

Masako Yasutake, Yasuo Koyama, Kenji Yoshimura, Kosho Shudo. 1997. Fixed-collocations and Their Permissible Variants. The Proc. of the 3rd Annual Meeting of ANLP: 449-452. (in Japanese)

Dan Sperber and Deirdre Wilson. 1995. Relevance: Communication and Cognition: Second Edition, Blackwell.