MWEs as Non-propositional Content Indicators

Kosho Shudo, Toshifumi Tanabe, Masahito Takahashi† and Kenji Yoshimura

Fukuoka University
8-19-1, Nanakuma, Fukuoka,
814-0180 JAPAN

†Kurume Institute of Technology
66-2228, Kamitsu, Kurume,
830-0052 JAPAN

{shudo,tanabe,yosimura}@tl.fukuoka-u.ac.jp taka@cc.kurume-it.ac.jp

Abstract

We report that a proper employment of MWEs concerned enables us to put forth a tractable framework, which is based on a multiple nesting of semantic operations, for the processing of non-inferential, Non-propositional Contents (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, the relationship between *indirect* (Searle, 1975) and *direct* speech, and equations peculiar to modal logic and its family (Mally, 1926; Prior, 1967) are treated in the similarity paradigm.

1 Introduction

While proper treatment of the Propositional Content (PC) of a sentence is undoubtedly important in natural language processing (NLP), the Non-propositional Content (NPC) also plays a critical role in tasks such as discourse understanding, dialogue modeling, detecting speaker's intension. We refer generically to the information which is provided by auxiliaries, adverbs, sentence-final 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), has been discussed in semantics or pragmatics, it lies beyond the state-of-the-art technology of NLP. Besides, no systematic attempt to connect linguistic forms in the sentence with the non-inferential NPCs 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, particles and their various alternative multiword expressions. We have extracted extensively these expressions 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 practical NLP tasks.

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 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 introduce a framework for the treatment of NPC. A set of primitive functions to compose NPC is explained in Section 4. In Section 5, first, the relationship between the framework and Japanese syntax, and second, methods to identify NPCs of Japanese sentences and to apply them to a translation task are described. In Section 6, we formalize the similarity among NPCs within the framework. In Section 7, we present conclusions and comment on future work.

2 Background MWE Resources

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

Second ACL Workshop on Multiword Expressions: Integrating Processing, July 2004, pp. 32-39

In this Section, we present an overview of our ongoing development of Japanese MWE resources. We have extracted multiword expressions that take 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 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 bone*)*'* is endowed with (1,0,0) and compositional, separable and statistically-bound expression, "ぐっすり·眠る *gussuri·nemuru*" *'sleep soundly',* with (0,0,1). A dot '·' denotes a conventional word-boundary, hereafter.
Fixed expressions, decomposable idioms, institutionalized phrases, syntactically-idiomatic phrases, light verb constructions discussed in (Sag et al., 2002) and proverbs might correspond roughly to the triplets, (1,1,0), (1,0,0), (0,0,1), (0,x,1), (0,x,1) and (1,1,1), respectively.
MWEs, whose number amounts to 64,800 at present, are classified by their overall, grammatical functions as follows. Examples with a triplet and the current number of expressions are also given in the following. Compound nouns and proper nouns are excluded in the present study.

Conceptual MWEs:

nominal<10,000>:"赤·の·他人 *aka·no·tanin*" (1,1,0) *'complete stranger* (lit. *red stranger*)*'*; "鶴·の·一声 *turu·no·hitokoe*" (1,1,0) *'the voice of authority* (lit. *one note of crane*)*'* ; etc.

verbal-nominal<1,700>:"もらい·泣き *morai·naki*" (1,1,0) *'weeping in sympathy* (lit. *received crying*)*'*; "ラッパ·飲み *rappa·nomi*"(1,1,0) *'drinking direct from the bottle* (lit. *trumpet drink*)*'*; etc.

verbal<34,000>: "かみ·締める *kami·simeru* "(1,1,0) *'chew well* (lit. *bite and fasten*)*'*; "煮·詰める *ni·tumeru*" (1,1,0) *'boil down* (lit. *boil and pack in*)*'*; etc.

adjectival<4,300>: "怒り·っぽい *okorip·poi* "(0,1,0) *'irritable* (lit. *anger-ish*)*'*; "注意·深い *chuui·bukai*" (1,1,0) *'careful* (lit. *deep in caution*)*'*; etc.

adjectival-nominal<2,000>:"一巻·の·終わり *ikkan·no·owari*" (1,1,0) *'the very end* (lit. *the end of a roll*)*'*; "筋書き·通り *sujigaki·doori*"(0,1,0) *'as just planned* (lit. *just as a plot*)*'*; etc.

adverbial<5,200>: "悪く·する·と *waruku·suru·to*" (1,1,0) *'if the worst happens* (lit. *if it worsens*)*'*; "うっとり·と *uttori·to*"(0,1,0)*'abstractedly'*; etc.

adnominal<2,600>: "他愛·の·無い *taai·no·nai*" (1,0,1)*'inconsiderable* (lit. *with no altruism*)*'*; "断固·たる *danko·taru*"(0,1,0) *'firm'*; etc.

connective<300>: "その·結果 *sono·kekka*" (1,1,0) *'consequently* (lit. *the result*)*'*; "それ·は·さて·おき *sore·ha·sate·oki*" (1,1,1) *'by the way* (lit. *setting it aside*)*'*; etc.

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

proverb-sentential-incomplete<900>: "病·は·気·から *yamai·ha·ki·kara*" (1,1,0) *'Fancy may kill or more.* (lit. *Illness is brought from one's feeling.*)*'*; "馬·の·耳·に·念仏 *uma·no·mimi·ni·nenbutu*" (1,1,1) *'A nod is as good as a wink to blind horse.* (lit. *buddhist's invocation to the ear of a horse*)*'* ; etc.

Functional MWEs:

relation-indicator(RI)<1,000>:"に·つい·て *ni·tui·te*" (1,1,0) *'about* (lit. *in touch with*)*'*; "に·よっ·て *ni·yot·te*" (1,1,0) *'by* (lit. *depending on*)*'*; "と·とも·に *to·tomo·ni*" (1,1,0) *'with* (lit. *accompanied with*)*'*; "に·おける *ni·okeru*" (1,1,0) *'in','on* (lit. *placed in*)*'*; etc.

NPCI<1,500>: See Section 4.

Nominals listed above are those marked with a triplet (1,1,x). We exclude compound nouns with (0,0,x) and proper nouns, whose number amounts to quite large, in this study. They should be treated in some other way in NLP. A treatment of those compound nouns for Japanese language processing is reported in (Miyazaki et al., 1993).
Formally, the triplet is expanded in the lexicon to a partly multi-valued 7-tuple (f_1, f_2, f_3, f_4, f_5, f_6, f_7). The augmented features are as follows;

f_4: grammatical class (shown above)
f_5: syntactical, original internal-structure

[1] At present f_1 and presumably f_2 will not be decided by any statistical method.

f_6: morphosyntactical variation: (m_1, m_2, ... , 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 words
m_4: particles insertable in between its 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 list of MWEs but designed as a resource proliferous to a total variety of idiosyncratic expressions. (Shudo et al., 1980, 1988; Shudo, 1989; Yasutake et al., 1997).
The present study focuses on a set of NPCIs and its relationship to the non-propositional structure of natural sentences. Some of our multiword NPCIs are treated in the general, rewriting framework for MT in (Shirai et al., 1993).

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 by three 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[2]. According to the occurrences of them, 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, in contrast, 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., $OBLIGATION_2$, $NEGATION_1$, PAST-TENSE, and NEG-EVAL, we can explain the Non-propositional Structure (NPS) of (2) as;

(4)PAST-TENSE [$NEGATION_1$ [$OBLIGATION_2$["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'*]].[3]

Here, a problem is that (4) is wrong for (1). In order to cope with this, while adopting a MWE, "べきで·なかっ·た *bekide·nakat·ta*" as a NPCI with a triplet (1,0,0) which has a composite NPF, NEG-EVAL[PAST-TENSE[x]][4], 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.[5] This is caused by its idiomaticity, i.e., by the difficulty in decomposing it into semantically consistent sub-forms.
Investigating a reasonably sized set of Japanese linguistic data, keeping the strategy exemplified above in mind, revealed that NPS of a natural Japanese sentence can be generally formulated as a nested functional form;

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

where S is a propositional, kernel sentence; M_i ($1 \leqq i \leqq n$), a NPF. In the following, we use the

[2] "べきだ *bekida*" and "ない *nai* " are inflected as "べきで *bekide*" and "なかっ *nakat*", respectively, in (1).

[3] We use lower-suffixes to distinguish NPFs by the subtle differences in meaning, degree, etc.

[4] 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.

[5] 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]].

notation for a composite function, $M_n \circ M_{n-1} \ldots \circ M_2 \circ M_1$, where $M_n \circ M_{n-1} \ldots \circ M_2 \circ M_1[S] = M_n[M_{n-1} \ldots [M_2[M_1[S]]] \ldots]$.

4 NPCIs, NPFs

We have settled a set of 150 basic NPFs by classifying 1,500 NPCIs which had been extracted from the large-scale data. After manually extracting them, the data has been continuously checked and updated by comparing with various dictionaries and linguistic literature such as (Morita et al.,1989).
They are subclassified as follows, though the boundaries between subclasses are partly subtle. It should be noted that some NPCIs are semantically ambiguous, being included in different subclasses below. Examples of NPCIs and the number of NPFs are given in brackets, in the following list.

F_1:polarity <3>:
NEGATION$_1$(“ない *nai*” *‘not’* ; “の·で·は·ない *no·de·ha·nai*”(1,0,0) *‘not’* ; etc.),
NEGATION$_2$(“と·いう·訳·で·は·ない *to·iu·wake·de·ha·nai*”(1,0,0) *‘not’* ; etc.),etc.

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

F_3:aspect-observational <9>:
IMMEDI-AFT-TERMINATING (“た·ところ·だ *ta·tokoro·da*”(1,1,0) *‘have just* V-en*’*; “た·ばかり·の·ところ·だ *ta·bakari·no·tokoro·da*” (1,1,0) *‘have just* V-en*’*; etc.),
IMMEDI-BEF-BEGINING(“う·と·し·て·いる *u·to·si·te·iru*” (1,0,0) *‘*be *about to’*; “よう·と·し·て·いる *you·to·si·te·iru*” (1,0,0) *‘*be *about to’* ; etc.),
PROGRESSING(“て·いる *te·iru*” (1,0,0) *‘*be V-ing*’*; “つつ·ある *tutu·aru*”(1,1,0) *‘*be V-ing*’*; etc.), etc.

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…*’*),
PAS-SUFFERING(“れる *reru*” *‘have*…V-en*’* ; “られる *rareru*” *‘have*…V-en*’*; etc.),
PAS-BENE-TAKING$_1$ (“てもらう *te·morau*”(1,0,0)*‘ask* …V*’*; “て·いただく *te·itadaku*” (1,0,0) *‘ask*… V*’*; etc.),
BENE-TAKING(“て·くれる *te·kureru*” (1,0,0) *‘*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 *te·miru*” (1,0,0) *‘try to’* ; etc.),etc.

F_8:modality <60>:
NEG-EVAL(“べき·で·ない *beki·de·nai*” (1,0,0) *‘should not’* ; “の·は·よく·ない *no·ha·yoku·nai*” (1,0,0) *‘should not’* ; etc.),
OBLIGATION$_2$(“必要·が·ある *hituyou·ga·aru* ” (1,0,0) *‘need’*, “べきだ *bekida*” *‘should’*, etc.),
OBLIGATION$_1$(“なければ·ならない *nakere·ba·nara·nai*” (1,1,1) *‘have to’* ; etc.),
PROHIBITION(“て·は·ならない *te·ha·nara·nai*”(1,0,1) *‘should not’*, etc.),
CAPABILITY(“得る *uru*” ; “こと·が·できる *koto·ga·dekiru*”(1,0,0) *‘*be *able to’*; etc.),
GUESS$_1$(“う *u*” *‘will’*), etc.

F_9:illocutionary-act <28>:
IMPERATIVE(imperative-form of verb *‘*imperative form*’*),
INTERROGATIVE (“か *ka*” ‘interrogative form*’*; “の·か *no·ka*”(1,1,0) ‘interrogative form*’* ; etc.), PROHIBITIVE(“な *na*” *‘Don’t... ’*), PERMISSIVE(“て·よい (1,1,0) *te·yoi*” *‘You may...’* ; “て·も·かまわ·ない *te·mo·kamawa·nai*”(1,0,0) *‘You may...’* ; etc.),
REQUESTING(“てください *te·kure*”(1,1,0) *‘Please...’*; “て·ほしい *te·hosii*”(1,1,0) *‘I want you to...’* ; etc.), etc.

5 Treatment of NPSs

5.1 Sentence-final Structure in Japanese

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

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

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 use predicative parts, $PRED \cdot m_1 \cdot m_2 \cdot \cdots \cdot m_n$ instead of full sentences, for simplicity.
Our morphology model was developed so as to fit for the general semantic processing, adopting MWEs. It is a probabilistic finite automaton with

150 states that prescribes minutely the internal structure of each BP and the predicative part. We leave its detail to (Shudo et al., 1980).

5.2 Identifying NPS

Based on our morphological analyzer, we have developed a segmenter (SEG) that segments the input predicative part into a PRED and each NPCI, and a NPS-constructor (NPSC) that constructs NPSs. For example, an input;

(9) "読まなければならないだろう *yomanakerebanaranaidarou*" *'will have to read'*

is first segmented into

(10) "読ま/·なけれ·ば·なら·ない/·だろ·う *yoma/·nakere·ba·nara·nai/·daro·u* "

by SEG. Here, a slash '/' denotes a segment-boundary identified by SEG. Then, NPSC evaluates a function *nps* defined below.

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

where $M_i^k[\ldots M_i^2[M_i^1[x]]]$ is a NPF (if k =1) or a composite of NPFs (if k≧2) associated with m_i. Hence, the computation of *nps* for (10) is;

(12) *nps*("読ま/·なけれ·ば·なら·ない/·だろ·う *yoma/·nakere·ba·nara·nai/·daro·u*")
=$GUESS_2$[*nps*("読ま/·なけれ·ば·なら·ない *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 $GUESS_2$ and $OBLIGATION_1$ are associated with "だろ·う *daro·u*" *'will'* and "なけれ·ば·なら·ない *nakere·ba·nara·nai* " *'have to'*, respectively. In order to examine the adequacy and sufficiency of NPFs, we evaluated outputs of NPSC for 4,083 input predicative parts, which had been taken randomly as a test set from newspaper articles and segmented by SEG. It produced a recall of 97.4% and a precision, 41.8%. The score of the recall seems to imply the sufficiency of the set of NPFs and NPCIs. Relatively low score of the precision is due to the system's over-generation caused by the semantic ambiguities of NPCIs. Among various measures to be taken, firstly, semantic constraints to control the composition operation '◦' may be effective to produce a better precision. The complete disambiguation measure is left to future work.

5.3 Application to J/E Machine Translation

We introduce here another experimental system, referred to as ENGL, whose input is the NPS of a sentence and whose output is its English forms, to demonstrate the usefulness of our formalism. ENGL simply realizes NPFs within English syntax. We assumed each NPF for English could be accomplished by applying rewriting rules of two types; i) $V \rightarrow x \cdot V_v \cdot y$ and ii) $S \rightarrow x \cdot S_v \cdot y$, 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.
Basically, a single rewriting rule is applied for a single NPF. However, occasionally, a NPF requires several rules to be applied successively. Also we may have no NPCI corresponding to a given NPF within the target language. For example, POLITENESS, which is common in colloquial Japanese, has mostly no NPCI in English.
For example, the computation for (12) is

$GUESS_2$ [$OBLIGATION_1$ ["読む *yomu*"]]
= $GUESS_2$ [$OBLIGATION_1$ [*read*]]
= $GUESS_2$ [*have to · read*]
=*will · have to · read,*

where the rewriting rules associated with $NECESSITY_1$ and $GUESS_2$ are $V \rightarrow$ *have to* $\cdot V_{root}$ and $V \rightarrow$ *will* $\cdot V_{root}$, respectively.

We give four more I/O examples In (14), the instantaneous aspect of *aruki · hajimeru* ; *begin walk-ing* excludes the possibility of the interpretations, $PROGRESSING_1$, $PROGRESSING_2$ and STATE-OF-THINGS of *teiru*, which remain in (13) or (15). This is because the system deals with concatenatability rules based on aspect features of the predicate. (ENGL simply denotes the verb's inflected form by *-ed*, *-en*, etc.)

(13) *nps*("学ん/·で·いる ; *manan/·de·iru*")
$=_1$ $PROGRESSING_1$[*study*]= *be study-ing*,
$=_2$ $PROGRESSING_2$[*study*]= *have be-en study-ing*,
$=_3$ $COMPLETED_1$[*study*]=*have study-en*

(14)*nps*("歩き/·始め·て·いる ; *aruki/·hajime·te·iru*")
=$COMPLETED_1$[INCHOATIVE[*walk*]]= *have begin-en walk-ing*

(15)*nps*("愛し/·て·いる ; *aisi/·te·iru*")
= STATE-OF-THINGS[*love*] = *love*

(16)*nps*(“動かし/·て·み/·て·も·よい/·の·でしょ/·う /· か ; *ugokasi/·te·mi/·te·mo·yoi/·no·desho/·u/·ka*”)
=NTERROGATIVE[GUESS$_1$[DECLARATION [PERMISSIVE[TRIAL[*move*]]]]]
= *Will it be allowed that...try to move....?*

A small-scale experiment, for 300 NPSs extracted from sentences in technical papers has shown that ENGL produced a precision of 86% and a recall, 80%. While these relatively high scores implies the fundamental validity of the NPF framework, more extensive tests will be required to make more reliable evaluation for the general domain, since technical papers tend to have less-complicated NPFs. In addition, further correction and refinement of synthesis rules for English will be necessary to obtain higher scores.

6 Similarity between NPSs

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

6.1 Logical Rules

First, we discuss, here, the logical similarity relation, $\rightleftarrows \subset((\cup \mathbf{F_i})^*)^2$, $(1 \leqq i \leqq 8)$, which seems crutial for NLP tasks such as text retrieval or paraphrasing.[6] We prefer the term, ‘similarity’ to ‘equivalence’ here since it should be based on truth values taken in ‘most situations’, or in some ‘similar’ worlds.[7]
There are basic rules such as;

(17) NEG-EVAL◦ NEGATION$_1$ $\rightleftarrows$OBLIGATION$_2$
(18) NEGATION$_1$◦PERMISSION $\rightleftarrows$PROHIBITION,
(19) NEGATION$_{1,2}$◦NEGATION$_{1,2}$ $\rightleftarrows$ λ (identity function),
(20) $N \circ \lambda \rightleftarrows \lambda \circ N \rightleftarrows N$ for $^{\forall}N \in (\cup F_i)^*$, $(1 \leqq i \leqq 8)$,
(21) POLITENESS $\rightleftarrows$ λ,

(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 set of logically notable rules. For example, from the observation that “働い/·て·ばかり·いる/·訳·で·は·ない *hatarai/·te·bakari·iru/·wake·de·ha·nai*” *‘do not always work’* is similar to “働か·ない/·時·が·ある *hataraka·nai/·toki·ga·aru*” *‘It happens occasionally that…do not work’* the following rule will be induced;

(22) NEGATION$_2$◦HIGHEST-FREQUENCY $\rightleftarrows$LOW-FREQUENCY◦NEGATION$_1$.

Also, “働か/·なく/·て·も·よい *hataraka/·naku/·te·mo·yoi*” ‘*need not work’*; *‘It is allowed that...do not work…’* and “働か/·なけれ·ば·なら·ない/·事·は·ない *hataraka/·nakere·ba·nara·nai/·koto·ha·nai*” *‘It is not obligatory that …work…’*, will induce a rule;

(23) PERMISSION◦NEGATION$_1$ $\rightleftarrows$NEGATION$_1$◦OBLIGATION.

These rules can be generalized as (24), (24’) by introducing a ‘duality’ function, $\boldsymbol{d}$ defined below;

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

(24) NEGATION$_{1,2}$◦$\boldsymbol{M} \rightleftarrows \boldsymbol{d}(\boldsymbol{M})$ ◦NEGATION$_{1,2}$,

(24’) $\boldsymbol{M}$ $\rightleftarrows$NEGATION$_{1,2}$◦ $\boldsymbol{d}(\boldsymbol{M})$ ◦NEGATION$_{1,2}$.

We show two more examples;

(22’) HIGHEST-FREQUENCY◦NEGATION$_1$ $\rightleftarrows$NEGATION$_2$◦LOW-FREQUENCY.
nps(“働か/·ない·で·ばかり·いる *hataraka/·nai·de·bakari·iru*” *‘It is always that…do not work…’*) $\rightleftarrows$*nps*(“働く/·こと·が·ある/·と·は·言え·ない *hataraku/·koto·ga·aru/·to·ha·ie·nai*” *‘It does not happen that…sometimes work…’*).

(23’) OBLIGATION◦NEGATION$_1$ $\rightleftarrows$NEGATION$_1$◦PERMISSION.
nps(“働い/·て·は·なら·ない *hatarai/·te·ha·nara·nai*” *‘must not work’*) $\rightleftarrows$*nps*(“働い/·て·よい/·と·いう·事·は·ない *hatarai/·te·yoi/·to·iu·koto·ha·nai*” *‘It is not*

[6] While the NPF in $\mathbf{F}_i$, $(1 \leqq i \leqq 7)$ 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.

[7] But we do not enter further theoretical arguments here.

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 Pragmatic Rules

The similarity relation among the speaker's attitude or intention toward the hearer is defined as a set, $\backsim \subset \{ (a,b) \mid a,b \in (\mathbf{F}_1 \cup \mathbf{F}_2 \ldots \cup \mathbf{F}_9)^* \wedge ((^{\exists}i, 1 \leqq i \leqq 1 \wedge f_i \in \mathbf{F_9}) \vee (^{\exists}j, \ 1 \leqq j \leqq m \wedge g_j \in \mathbf{F}_9))$, 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,
  $\backsim$POLITENESS$\circ$IMPERATIVE,
  $\backsim$INTERROGATIVE$\circ$NEGATION$_1$
    $\circ$BENE-TAKING,
  $\backsim$INTERROGATIVE$\circ$NEGATION$_1$
    $\circ$CAPABILITY$\circ$PASS-BENE-TAKING,
  $\backsim$DESIRING$\circ$PASS-BENE-TAKING,
  $\backsim$DESIRING$\circ$PASSIVE.
nps("見/·て·くれ *mi/·te·kure*" *'Look at …'*),
  $\backsim$ *nps*("見る/·こと·が·出来る/·ない/·か *miru/·koto·ga·deki/·nai/·ka*" *'Can't you look at …?'*),
  $\backsim$*nps*("見る/·こと·が·出来る/·か *miru/·koto·ga·dekiru/·ka*" *'Can you look at …?'*),
  $\backsim$*nps*("見/·なさい *mi/·nasai*" *'Please look at …'*),
  $\backsim$ *nps*("見/·て·くれ/·ない/·か *mi/·te·kure/·nai/·ka*" *'Don't you look at … for me …?'*),
  $\backsim$ *nps*("見/·て·もら/·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/·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 …'*).

With respect to prohibition, invitation, permission and assertion, we 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 *hait/·te·ha·nara·nai*" *'You must not enter…'*) ,
  $\backsim$ *nps*("入る/·事·が·出来/·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 *ki/·te·yoi*" *'You may wear…'*)
  $\backsim$ *nps*("着る/·こと·が·出来る *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 ...? '*).

We have obtained approximately 30 pragmatic rules 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 Conclusions

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 here 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 will surely be further broadened by investigating non-final parts of Japanese sentences. This research should 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 frameworks presented here could hopefully provide tools for those comparative studies.

References

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

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.

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

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. 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)