

# When nouns need co-arguments

## A case study of semantically unsaturated nouns

Kow KURODA

Masaki MURATA

Kentaro TORISAWA

National Institute of Information and Communications Technology (NICT), Japan

3-5 Hikari-dai, Seika-cho, Sooraku-gun, Kyoto, 619-, Japan

### 1 Introduction<sup>1)</sup>

Recently, we finished the manual cleaning of approximately 67,000 Japanese hypernym hierarchies paired with roughly 900,000 hyponyms. The original data, comprised of roughly 2,400,000 hypernym-hyponym pairs, were automatically acquired from the Japanese Wikipedia (Sumida et al., 2008).<sup>2)</sup> The data we processed resemble the following:<sup>3)</sup>

- (1) a.  $h_1$ : singer,  $h_2$ : rock singer,  $h_3$ : British rock singer,  $h_4$ : famous British rock singer;  $I$ : Peter Gabriel
- b.  $h_1$ : member,  $h_2$ : former member,  $h_3$ : former member of Floyd,  $h_4$ : former member of Pink Floyd;  $I$ : Syd Barrett

Pairs ( $H$ ;  $I$ ), where  $H = h_1, \dots, h_n$ , are automatically generated from such pairs ( $h_{\max}$ ;  $I$ ). We refer to  $H$  as the “hyponym path” for  $I$ ,<sup>4)</sup> and to units like  $h_1, h_2, \dots, h_n$  as the “path elements” of  $H$ .

A hypernym path may contain: (i) bare nouns (e.g., *singer*), (ii) modified nouns (e.g., *famous British rock singer*, *former member*), or (iii) noun phrases. Such a path is constructed by automatically removing modifiers from  $h_n$  one by one. This operation is not error-free. Manual cleaning was performed to discard unacceptable units like *for-*

<sup>1)</sup>This paper benefited from the comments by three anonymous reviewers. Any remaining errors, however, are the responsibility of the authors—mainly of the first.

<sup>2)</sup>The data we processed are special because nearly 2/3 of the hyponyms are proper names. These data are expected to complement such traditional thesauri as WordNet (Fellbaum, 1998) in which upper ontologies are specified. We previously linked the roots of the hypernym hierarchies thus cleaned to the leaves of WordNet-Ja (WN-Ja) (Bond et al., 2009). Currently, 70% of the roots of the hypernym hierarchies are linked to WN-Ja (the coverage was only 8% in the original ( $h_n, I$ ) pairs).

<sup>3)</sup>We worked on Japanese data, but we present the English counterparts in this paper.

<sup>4)</sup>For both practical and theoretical purposes, we did not distinguish “instance-of” and “subclass-of” relations.

*mer member of Floyd*,  $h_4$  of (1b) yielded in the automatic generation.<sup>5)</sup>

It soon became apparent that checking for the conventionality alone was unworkable. We also needed a systematic treatment of units like (*former*) *member*, nouns that cannot be used independently and seem to require “arguments” at least semantically. Terms like *singer* make good hypernyms, but not terms like (*former*) *member* due to their semantic unsaturatedness. By contrast, (*former*) *member of Pink Floyd* creates good hypernyms; yet no shorter terms can be accepted as good hypernyms. Thus, semantically unsaturated nouns probably do not make good hypernyms. If correct, however, it poses another question: **exactly what terms behave semantically unsaturated?** The answer, which is far from obvious, motivated us to formulate and develop the descriptive model presented below.

The paper is organized as follows: In §2, we examine crucial cases of argument-taking nouns of (i) deverbal nouns, (ii) nondeverbal relational nouns, and (iii) nondeverbal quasi-relational nouns, and try to reinterpret their behavior in terms of their “frame-evoking” property, adopting a useful notion from FrameNet (Baker et al., 1998; Fontenelle, 2003). In the attempt, we introduce the notion of “co-arguments” that helps us to identify a class of “(semantically) unsaturated nouns” as a generalization over the relevant classes, and then we suggest a method for automatically measuring unsaturatedness of nouns using information about their “qualia structures” (Pustejovsky, 1995). In §3, we describe a few interesting properties of unsaturated nouns in this generalized sense.

<sup>5)</sup>We did not use Named Entity tagger to filter them out.

## 2 Description of Phenomena

### 2.1 Argument structures of nouns

#### 2.1.1 Case of deverbal nouns

Let us begin with simpler cases like *singer*, *inventor*, *creation*, and *invention*. These are quite frequent deverbal hypernyms in the data: *singer.n* is derived from *sing.v*, *inventor.n* is from *invent.v*, and *creation.n* is derived from *create.v*. Due to their nature, these nouns clearly have argument structures, inherited from base verbs, which are realized in nominal forms in different ways:

- (2) a. The *inventor* of WWW is Tim Berners-Lee.
- b. Linux is a *creation* { i. by; ii. of } Linus Torvalds.

#### 2.1.2 Case of nondeverbal nouns

Are deverbal nouns the only nouns with argument structures? It seems not. Consider cases like *author*, *masterpiece*, and *song*, which are not deverbal, but they seem to have argument structures like the deverbal nouns presented in (2).

- (3) a. The *author* of “Gargantua and Pantagruel” is François Rabelais.
- b. “The Creation” is a *masterpiece* of Joseph Haydn.

The parallelism between the cases in (2) and (3) is obvious, but it is hard if not impossible to claim that nouns like the ones in (3) are derived from verbs. For example, *author* is not deverbal, even if ending with agentivity-marking morpheme *-or*.<sup>6)</sup>

#### 2.1.3 Case of so-called “relational” nouns

The relevant property of nouns like *author* is related to the notion of **relational nouns** exemplified by *sister*, *mother*, and *father* which arguably have their own arguments, despite not being derived from any verbs. Many linguists have concentrated on such nouns of this class (de Bruin and Scha, 1988; Nishiyama, 1990; Partee and Borschev, 2003) mainly due to the peculiarities in their semantics. Let us cite a relevant definition of relational nouns:

- (4) Relational nouns are semantically unsaturated. They are normally used in combination with an implicit or explicit *argument*: *John’s brother*. The argument of a relational noun, if overtly realized in the sentence, is

connected to the nouns by means of a relation-denoting lexical element: the verb *have* or one of its semantic equivalents (the genitive and the prepositions *of* and *with*): *John has a sister*, *John’s sister*, *a sister of John’s*, *a boy with a sister*.

The definition of relational nouns cited from de Bruin and Scha (1988) represents the notion of them widely accepted in linguistics.

A similar notion is adopted in cognitive psychology (Gentner, 2005; Gentner and Kurtz, 2005), but there seems to be a subtle difference. Most linguists would be reluctant to admit that *masterpiece*, *letter*, *victim*, *part*, *member*, and (*vice-*)*president*, are relational nouns, yet cognitive psychologists would not; they do not seem to hesitate to consider them as relational nouns, hypothesizing that relational nouns are any nouns that denote (components of) **relational categories** (Gentner, 2005; Gentner and Kurtz, 2005). It is not trivial to tell what nouns count as relational and what nouns as not.

Let us pose a crucial question, Is “*former member*” in (1b) a relational noun? It is likely, but confirmation is not trivial. Differently put, it is hard, if not impossible, to differentiate true relational nouns from “quasi-relational” nouns such as (*former*) *member*, *masterpiece* and *letter*. The reason is two-fold: First, *former* adds unsaturatedness to *member*, which itself is unsaturated, suggesting *former* has unsaturatedness of its own. It is not clear at all how they interact, however. Second, as far as we know, nobody has ever tried to exhaustively list relational nouns for the lack of reliable criteria to identify relational nouns.

Under this, this paper proposes a set of prerequisites for effective identification of relational nouns. Before starting it, however, let us first specify the conceptual problem that needs be solved. If non-deverbal nouns including relational nouns can have argument structures, what nouns **cannot**? Does *masterpiece* in (3), for example, have an argument structure or not? The answer is not obvious. More specifically, are all nouns with argument structures relational nouns? To answer this successfully, we need to extend the notion of relational nouns to (**semantically**) **unsaturated nouns**.

#### 2.1.4 Co-argument structures of nouns

Note that nouns like *inventor*, *invention*, *author*, and *masterpiece* are either names for **frame elements** (FEs) in the sense of Berkeley FrameNet

<sup>6)</sup>Etymological consideration is irrelevant here.

(BFN) (Baker et al., 1998; Fontenelle, 2003) or names for **representative instances** for relevant FEs. In fact, *inventor* and *invention* name FEs  $\langle \text{Cognizer} \rangle$  and  $\langle \text{Invention} \rangle$  of  $\langle\langle \text{Invention} \rangle\rangle$ <sup>7)</sup> ( $\langle\langle X \rangle\rangle$  denotes a frame, and  $\langle X \rangle$  denotes a frame element hereafter). Similarly, *author* names an FE  $\langle \text{Author} \rangle$  of  $\langle\langle \text{Text\_creation} \rangle\rangle$ <sup>8)</sup>. Also, *inventor* and *invention* are **lexical units** (LUs) for  $\langle\langle \text{Invention} \rangle\rangle$  in that they “evoke” the frame. *author* is a lexical unit for  $\langle\langle \text{Text\_creation} \rangle\rangle$ .

The situation is different for nouns like *masterpiece*; *master* is a lexical unit for  $\langle\langle \text{Expertise} \rangle\rangle$ <sup>9)</sup>, but it is obscure how it is related to  $\langle\langle \text{Text\_creation} \rangle\rangle$ . It is also unclear how the semantics of *piece* is incorporated, making it difficult to recognize *masterpiece* as a name for a specific FE.

A promising solution would be to assume that nouns  $N$  evoke frame  $F$  strongly enough if they have strong instantiation biases for core FEs of  $F$  even if  $N$  does not name an FE. Thus, *masterpiece*, for one, is biased for the instance of  $\langle \text{Created\_entity} \rangle$  of  $\langle\langle \text{Creating} \rangle\rangle$ .<sup>10)</sup>

Readers should already be aware that what is relevant here is the implicit **co-argument structure** of noun  $N$ , either  $N$  is deverbal or not. Noun  $\alpha$  has co-arguments  $\{\beta_1, \dots, \beta_n\}$  relative to predicate  $P$  if and only if  $\alpha$  and  $\beta_1, \dots, \beta_n$  are arguments of  $P$  and  $P$  is a LU for frame  $F$ . For example,  $\langle \text{Creator} \rangle$  and  $\langle \text{Created\_entity} \rangle$  are **co-FEs** relative to the  $\langle\langle \text{Creating} \rangle\rangle$  frame. Other (core) FEs such as  $\langle \text{Cause} \rangle$  are all co-FEs of each other. The notion of co-FEs is useful, but it depends on a specific theory, FrameNet. We looked for a more theory-neutral term, and adopted “co-arguments.”

## 2.2 Semantically unsaturated nouns

Based on the set of issues described so far, we define the “semantically unsaturated nouns” as a cover term for the relevant classes of nouns:

- (5) **(Semantically) unsaturated nouns**<sup>11)</sup> are nouns that has either (proper) arguments

<sup>7)</sup><http://framenet.icsi.berkeley.edu/fnReports/displayReport.php?frame=Invention>

<sup>8)</sup>[http://framenet.icsi.berkeley.edu/fnReports/displayReport.php?frame=Text\\_creation](http://framenet.icsi.berkeley.edu/fnReports/displayReport.php?frame=Text_creation)

<sup>9)</sup><http://framenet.icsi.berkeley.edu/fnReports/displayReport.php?frame=Expertise>

<sup>10)</sup>Regrettably, the inheritance link from  $\langle\langle \text{Creating} \rangle\rangle$  to  $\langle\langle \text{Text\_creation} \rangle\rangle$  is missing in the current BFN database.

<sup>11)</sup>Note that the term “unsaturated nouns” was first coined by Yuji Nishiyama in his work (Nishiyama, 1990) to charac-

or co-arguments. This class of nouns include: (i) nouns derived from predicates such as verbs and adjectivals, (ii) relational nouns, and (iii) (quasi-relational) nouns with (strong) co-arguments.

The definition in (5) is rather informal. We will give a more formal definition under Formal Concept Analysis (FCA) (Ganter et al., 2005) in §3. But let us address an unsolved issue before proceeding. Note that the definition in (5) relies on the assumption that there is a procedure with which we can identify a given noun’s arguments and co-arguments. Admittedly, effective identification of co-arguments is hard to achieve. Let us specify our idea for it, though it is not implemented yet.

### 2.2.1 Identifying co-arguments through qualia structure

A closer examination of nouns with “strong” co-arguments (e.g., *masterpiece* and *letter*) suggests that they tend to have “clear” and “distinct” qualia structures (Pustejovsky, 1995). We say that nouns have clear qualia structures if they have well-defined purposes (clear and determined telic roles) and well-understood origins (clear and determined agentive roles).

It is not a coincidence that clearly unsaturated nouns tend to have distinct qualia structures, although it is unclear if the aspects of unsaturatedness are all derivatives of the qualia structure. But the contrary does not seem to be true: *Beer*, for example, has a clear qualia structure that defines *brewer* or *brewery* bearing agentive role, but it is not clear whether this implies that *beer* is an unsaturated noun. For evidence, the *by*-phrase in *beer by Budweiser* is more likely to be treated as a modifier than as an argument. But the notion of co-argument allows us to say that *by*-phrase is a co-argument of *beer*.

This implies that so-called “modifiers” on a noun  $N$  fall into three types: (i)  $N$ ’s arguments,  $N$ ’s co-arguments (often licensed via qualia structure) and (iii)  $N$ ’s pure modifiers. Typical example of pure modifiers are specifiers of time and place. This provides a new insight into the classification of modifiers.

terize the behavior of Japanese relational nouns in the sense of de Bruin and Scha (1988) presented in (4). For the lack of a better term, we redefined Nishiyama’s term to denote a larger set of nouns and hope this does not mar his contribution to Japanese theoretical linguistics.



	A	B	C	D	E	F	G	H	I	J	K	L	M
	takes-mo...	mark-arg...	takes-no...	is-a-verb	derived-f...	allow-mo...	denote-in...	takes-onl...	inflects	needs-su...	needs-aux	is-a-noun	
predicate													
verbal SA-HEN noun		X	X		X		X			X		X	X
relational noun		X							X				X
unsaturated noun		X					X						X
non verbal SA-HEN noun		X			X								X
non deverbal noun		X	X				X			X	X		X
verbal noun		X				X							X
multiplex relational predicate		X					X						X
relational predicate		X											X
adjective		X	X						X				X
adjectival noun		X	X						X			X	X
proper noun				X				X					X
saturated noun				X									X

Figure 2: Attribute-value matrix used to build the FCA in Fig. 1

is the most straightforward explanation of why nouns **can** have arguments and co-arguments.

### 3.1.1 Note on Japanese verb morphology

As mentioned above, we do expect the classification to be naturally applicable to languages other than Japanese, but there is a peculiar fact about Japanese morphosyntax that complicates it. We take a brief look into it.

In the analysis in Fig. 1, noun *X* is called a “SA-HEN noun” (サ変名詞 in Japanese) if *X* is comprised of a verb of the form *X-suru*, with no case marker intervening between *X* and *suru*, a kind of light verb.<sup>12)</sup> Let us explain the relevant facts using an example. In (7), nominative NP *sakusha-ga* is marked as ⟨Creator⟩, and accusative NP *sakuhin-wo* as ⟨Created\_entity⟩ of ⟨⟨Creating⟩⟩.

#### (7) Clausal form:

*Sakusha-ga sakuhin-wo sousaku-suru.*  
author-NOM work-ACC create-does  
“An author creates (a piece of) work.”

#### (8) Nominal(ized) forms of (7) meaning “creation of work by an author ...”

- a. *Sakusha-no sakuhin-no sousaku(-ga ...).*  
author-OF work-OF creation(-NOM ...)
- b. *Sakusha-niyoru sakuhin-no sousaku(-ga ...).*  
author-BY work-OF creation(-NOM ...)

<sup>12)</sup>Kageyama (1993) called *X* a “verbal noun.” This terminology is quite influential, both in and out of Japan, in the tradition of Generative Grammar (GG), but we decided not to use it because it is less accepted outside the narrow circle of GG; most researchers in Natural Language Processing (NLP) of Japanese, who know what SA-HEN nouns are, would not know what verbal nouns are.

Also, the proper characterization of SA-HEN nouns remains somewhat controversial. It is not clear whether they are zero-derived from corresponding verbs that have *-suru* underlyingly. Our position is that those nouns are underlyingly underspecified for the verb/noun distinction, and *-suru* is required when overt inflection is needed. This analysis is compatible with the lattice diagrammed in Fig. 1.

In this case, *X=sousaku*. In (7), *sousaku-suru* is a verb. In (8), *sousaku* is an unsaturated noun that require arguments semantically, *sakusha* and *sakuhin*, marked for adjuncts (by *niyuru* = “by,” *no* = “of”). (9) illustrates the English counterparts of the Japanese classes of (a) deverbals, (b) SA-HEN nouns, and (c) non-deverbal, non-SA-HEN nouns:

- (9) a. deverbal nouns (A): *invention.n, inventor.n, inventory.n* (from *invent.v*), *creation.n, creator.n* (from *create.v*)
- b. SA-HEN nouns (B):<sup>13)</sup> *walk.n* (= *walk.v*), *move.n* (= *move.v*)
- c. non-deverbal, non-SA-HEN nouns (C): *(counter)measure.n* (in *devise a (counter)measure against*), *antipathy.n* (\**antipathize.v*; cf. *sympathy.n, sympathize.v*)

### 3.1.2 Support verb constructions in Japanese

In Japanese, the syntax and semantics of nouns of type A and B nouns are regular, but this is not true of type C nouns: *taisaku.n* (“countermeasure”), *hankan.n* (“antipathy”) and *kugen.n* (“expostulation”) are of type C. They have the same kind of argument structures as type C nouns, which differ from type B nouns because they cannot be used as verbs by adding *-suru*: \**taisaku-suru*, \**hankan-suru*, and \**kugen-suru* are all unacceptable. Instead, they form support verb constructions, as shown in the following examples:

- (10) a. *X-ga Y-{ i. ni; ii. he(no) } taisaku-wo koujiru* (“X devise a (counter)measure against Y”)
- b. *X-ga Y-ni(taishite) hankan-wo { i. motsu; ii. idaku; iii. kanjiru }.* (“X { i. have; ii. feel } antipathy against Y”)

<sup>13)</sup>Nouns like *song* and *product* could be classified into this.

- c. *X-ga Z-nituite Y-ni(taishite) kugen-wo teisuru* (“X raise difficulties against Y over Z”)

where *taisaku*, *hankan* and *kugen* are used as accusative nouns (marked by *-wo*), and *kouji-ru.v* (“devise”), *motsu.v* (“have”), *idaku.v* (“hold”), *kanji-ru.v* (“feel”) and *tei-suru.v* (“show”) are support verbs for them.<sup>14)</sup>

### 3.2 Findings and Motivations

This section addresses a few important findings in the cleaning task that motivated the classification.

#### 3.2.1 Modifiers as sources of unsaturatedness

It is tempting to assume that unsaturatedness is a lexical property. But the reality of the data we processed suggests otherwise. We have seen many cases in which compound nouns acquire unsaturatedness by modifiers. For example, *city* is arguably a saturated noun, but *sister city* is unsaturated because *sister* adds unsaturatedness to it.

Note also that unsaturatedness accumulates through modification. Take “X is a *fellow disciple* of Y” for example. In this case, we can infer that both X and Y are *disciples* of Z, the same master. Interestingly, the unsaturatedness for *fellow* and *disciple* can co-exist, though the surface realization of Z seems to be suppressed.<sup>15)</sup> Note that *former* in (1b) behaves exactly like *fellow* here.

Unsaturatedness is composable, allowing the unsaturatedness of one noun to get “bridged” to another. In cases like *secretary of the Minister (of Foreign Affairs)*, unsaturatedness is reduced through variable-binding, because *secretary of X*, X is bound to *the Minister (of Y)*, and when Y is bound to *Foreign Affairs* (with the aid of *of*), it gets saturated; otherwise, it stays unsaturated.

#### 3.2.2 Unsaturated nouns are not rare

In this regards, it deserves a mention that linguists seem to tacitly assume that relational nouns are rare and exceptional, and that their set is closed. This assumption is far from well grounded because the sets of relational nouns and of nouns

<sup>14)</sup>In most constructions, support verbs are selected in complex ways influenced by subtle collocations, and achieving prediction with high precision is hard about what verbs can serve as support verbs for what type C nouns. This is probably true of similar cases in other languages.

<sup>15)</sup>Note here that a mutuality interpretation of relational nouns (Eschenbach, 1993) seems to have an interesting effect on the construction and interpretation of *sister cit(y)ies* and *fellow disciple(s)*.

with strong co-argument structures are often indistinguishable.

We state this based on our experience in the cleaning task explained in the Introduction. The task was performed based on classification guidelines that incorporated the definition of unsaturated nouns presented in (5). As a result, we obtained roughly 118,000 types of manually rated path elements. The ratios of saturated, unsaturated, dubious, and discarded elements were 57.4% 21.2%, 7.2%, and 14.2%, respectively. The net ratio of saturated against unsaturated was 73.0% against 27.0%, suggesting that nearly 1/5 of noun types are unsaturated. This does not seem to support the traditional view of unsaturated nouns.

#### 3.2.3 Degrees of unsaturatedness

The classification guidelines we prepared contain some heuristics to deal with the **degrees** of unsaturatedness that confused annotators. For example, the three Japanese nouns in (11) turned out to have different strengths of unsaturatedness and can be ordered in a specified way:

- (11) (X(sha)-no) sha-cho  
 (“president (of (company) X)”), *-sha* means “company”  
 < (X(bu)-no) bu-cho  
 (“manager (of section X)”)  
 < (X(kyoku)-no) kyoku-cho  
 (“director/chief (of department X)”)

Literally, (X-no) *cho* means “person of prime importance/head (of X).”

What is puzzling is that *shacho*, *bucho* and *kyokuchō*, all of which are relational and should therefore be equally unsaturated, show different degrees of unsaturatedness after human rating.

On a closer examination, the degrees of unsaturatedness, at least as measured through human judgments, correlated with distributional statistics: i) nouns with weaker word boundaries at the onset were judged to be more unsaturated; ii) nouns with stronger word boundaries at the onset were judged saturated even if they are theoretically unsaturated nouns: (X(*sha*)-no) *sha-cho* is one of such nouns.

Note that *shacho*, *bucho* and *kyokuchō* have different degrees of “wordhood.” Three statistics are relevant: First, the frequency order of the three terms is: *shacho* ≫ *bucho* > *kyokuchō*. It is likely that high frequency reduces the (feeling of) unsaturatedness of a term. Second, letting  $Y = \{ \textit{sha-cho}, \textit{bucho}, \textit{kyokuchō} \}$ , the varieties of *L* in the

context *LYR* follows the order of *shacho*  $\gg$  *bucho*  $>$  *kyokuchō*, irrespective of whether *L* ends with a case marker or not. Third, the relative frequencies of genitive marker *-no* against other markers follows the order of *shacho*  $\gg$  *bucho*  $>$  *kyokuchō*, given the context *XMYR* where *X* is a noun marked by marker *M* such as *-no*, *-ga*, *-de*, and *-wo*. The second and third are natural consequences of unsaturatedness. This observation suggests that *sha-cho* has the strongest independence.<sup>16)</sup>

Taking these properties into account, we hypothesized that potentially unsaturated nouns were sometimes wrongly classified as saturated nouns when they were frequent. Based on this, human raters were advised not to follow candid intuitive judgments during classification but to scrutinize the semantics of target nouns combined with potential arguments or co-arguments.

### 3.2.4 Measuring unsaturatedness

As an extension of this theorizing, we derived two different heuristic measures to differentiate unsaturated from saturated nouns. They are based on the distributional properties discussed in §3.2.3:

Heuristic 1: Unsaturated nouns appear less often at the beginning of a new NP especially when it begins a phrase.

Heuristic 2: Unsaturated nouns appear relatively more often after genitive markers than other case markers.

Japanese has postpositional particles for noun *N* such as: *N-ga* (nominative), *N-wo* (accusative), *N-ni* (dative, locative), *N-de* (instrumental, locative, manner), *N-kara* (ablative), *N-to* (committative), and *N-no* (genitive). Suppose we have a noun *N* to be checked for unsaturatedness. Heuristics 1 says that if *N* is unsaturated, the occurrence of a case marker is discouraged;<sup>17)</sup> if *N* is saturated, on the other hand, no such constraint is effective and

<sup>16)</sup>We are not sure if similar facts can be easily found in other languages: perhaps, this is a kind phenomenon is particular to agglutinative languages. The relevant fact is that the distinction of unbound morphemes (i.e. words) from bound morphemes is not categorical but stochastic in Japanese (and probably other agglutinative languages).

<sup>17)</sup>This is a consequence of a well-observed property of Japanese compound nouns: the occurrence of *-no* is discouraged in formal style of speech. For example, *tenki yohou* (“weather forecast”) can be paraphrased as *?tenki-no yohou* (“forecast of weather”). Nothing is wrong with this paraphrase from a semantic point of view, but the result sounds rather childish.

case-markers can occur more freely in front of *N*. This difference would yield the expected effect.

Heuristics 2 is related to a different aspect of Heuristics 1. Heuristics 2 argues that the relative frequency of genitive *-no* is larger for unsaturated nouns and smaller for saturated nouns, with other things being equal.

We originally planned to test the validity of the heuristics, but no further details are reported in this paper, since it is incomplete for the time being. But if either turns out to be valid, it would offer a quantitative definition of unsaturatedness, as we briefly touched on in §2.2.1.

### 3.3 Benefits of co-argument notion

Several projects (Meyers et al., 2004; Iida et al., 2007), some of which are ongoing, have investigated the annotation of argument-taking nouns in general domains. The typical annotation task marks the arguments of deverbal nouns like *inventor* and *invention* but it includes annotations for nondeverbal nouns like *masterpiece* and *letter*.<sup>18)</sup> Scrutiny of annotation guidelines (Meyers, 2007) reveals that no distinction is made between the proper arguments of deverbal nouns and the co-arguments of nondeverbal nouns. This does not look optimal.

Under the notion of co-arguments, *letter*, for example, can be marked for information of the kinds specified in (12),<sup>19)</sup> but such labels as  $\langle$ Author $\rangle$ ,  $\langle$ Addressee $\rangle$ ,  $\langle$ Medium $\rangle$ , and  $\langle$ Content $\rangle$  identify co-arguments rather than proper arguments:

(12) ...  $\langle$ Addressee: *he* $\rangle$  received a  $\langle$ Medium: *letter* $\rangle$  from  $\langle$ Author: *Nissan Motor* $\rangle$  stating  $\langle$ Content: *that the company had a concern over Nissan Computer's use of that domain name* $\rangle$ .<sup>20)</sup>

For some readers, the distinction of co-arguments from proper arguments probably resembles a mere terminological difference. But it is not, at least if we consider the qualia structure of a noun for its relevance. Given that **write**(*e*, *x*, *y*, *z*) and **inform**(*e*, *x*, *y*, *w*) are a *letter*'s agentive and telic components, respectively,  $x=Nissan$

<sup>18)</sup>All nouns exemplified here have entries in NomBank (release 1.0) [<http://nlp.cs.nyu.edu/meyers/NomBank.html>].

<sup>19)</sup>The original text was taken from the Web by Google.

<sup>20)</sup>We adopted BFN-style annotation due to its explicit reference to frame elements. In NomBank-style annotation, annotators need a mapping that mediate frame elements and such tags as Arg0, Arg1, ..., because reference to frame elements is indirect.

*Motor*,  $y=he$ ,  $z=letter$ , and  $w=$  “that the company had a concern ...”. It is not coincidence that FEs ⟨Author⟩, ⟨Addressee⟩, ⟨Medium⟩, and ⟨Content⟩ identify the functionalities of  $x$ ,  $y$ ,  $z$  and  $w$ .<sup>21)</sup> Note also that the configuration of these FEs is the same as BFN semantic role annotation except that the target is not *write*, a LU of ⟨⟨Creating\_text⟩⟩, but *letter*. Thus, it seems reasonable to hypothesize that the co-arguments of a noun  $N$  are licensed through  $N$ 's qualia structure, but it is not clear if this condition is necessary or sufficient against a large scale of data.

## 4 Conclusion

This paper proposed the notions of a noun's co-arguments differentiated from proper arguments and of semantically unsaturated nouns. We discussed some benefits from them. We hope they contribute to better understanding of nouns by unifying theories of semantic frames and qualia structures.

## References

- C. F. Baker, C. J. Fillmore, and J. B. Lowe. 1998. The Berkeley FrameNet Project. In *COLING-ACL 98, Montreal, Canada*, pages 86–90.
- F. Bond, H. Isahara, K. Uchimoto, T. Kuribayashi, and K. Kanzaki. 2009. Extending the Japanese WordNet. In *Proc. of the Annual Meeting of the Association of Natural Language Processing*, pages 80–83.
- P. Cimiano and J. Wenderoth. 2005. Automatically learning qualia structures from the web. In *Proc. of the ACL-SIGLEX Workshop on Deep Lexical Acquisition, Ann Arbor, Michigan, June 2005*, pages 28–37.
- J. de Bruin and R. Scha. 1988. The interpretation of relational nouns. In *Proc. of the Nth Annual Meeting of the Association for Computational Linguistics*.
- P. Eschenbach. 1993. Semantics of number. *Journal of Semantics*, 10(1):1–31.
- C. Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.
- T. Fontenelle, editor. 2003. *FrameNet and Frame Semantics*. Oxford University Press. A Special Issue of *International Journal of Lexicography*, 16 (3).
- B. Ganter, G. Stumme, and R. Wille, editors. 2005. *Formal Concept Analysis: Foundations and Applications*. Springer, Berlin/Heidelberg.
- D. Gentner and K. J. Kurtz. 2005. Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, and P. W. Wolff, editors, *Categorization Inside and Outside the Laboratory*, pages 151–175. APA.
- D. Gentner. 2005. The development of relational category knowledge. In L. Gershkoff-Stow and D. H. Rakison, editors, *Building Object Categories in Developmental Time*, pages 245–275. Hillsdale, NJ: Lawrence Erlbaum.
- R. Iida, M. Komachi, K. Inui, and Y. Matsumoto. 2007. Annotating a Japanese text corpus with predicate-argument and coreference relations. In *ACL Workshop Linguistic Annotation Workshop*, pages 132–139.
- T. Kageyama. 1993. *Grammar and Word Formation*. Hitsuji Publishing. [Original title: *Bunpoo to Gokeisei*].
- A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. The NomBank Project: An interim report. In A. Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 24–31, Boston, MA, May 2 - May 7.
- A. Meyers. 2007. Annotation guidelines for NomBank: Noun argument structure for PropBank.
- M. Murata, H. Isahara, and M. Nagao. 1999. Resolution of indirect anaphora in Japanese sentences using examples “ $x$ -no  $y$  ( $y$  of  $x$ )”. In *ACL'99 Workshop on Coreference and its Applications, Maryland, USA, June 22, 1999*, pages 31–38.
- Y. Nishiyama. 1990. On the “Kakiryori ha Hiroshimagahonba da” construction: Saturated and unsaturated noun phrases [in Japanese]. *Proc. of the Institute of Language and Culture, Keio University*, 22:169–188.
- B. H. Partee and V. Borschev. 2003. Genitives, relational nouns, and argument-modifier ambiguity. In E. Lang, C. Maienborn, and C. Fabricius-Hansen, editors, *Modifying Adjuncts, Interface Explorations 4*, pages 67–112. Mouton de Gruyter.
- J. Pustejovsky. 1995. *The Generative Lexicon*. MIT Press.
- R. Sasano, D. Kawahara, and S. Kurohashi. 2004. Automatic construction of nominal case frames and its application to indirect anaphora resolution. In *Proc. of the 20th International Conference on Computational Linguistics*, pages 1201–1207.
- A. Sumida, N. Yoshinaga, and K. Torisawa. 2008. Boosting precision and recall of hyponymy relation acquisition from hierarchical layouts in Wikipedia. In *Proc. of the 6th International Conference on Language Resources and Evaluation (LREC-2008)*.

<sup>21)</sup>Hence, *author*, *text*, *medium*, and *content* are names for semantic **roles** (or FEs) rather than semantic **types**.

- K. Torisawa. 2005. Automatic acquisition of expressions representing preparation and utilization of an object. In *Proc. of Recent Advances in Natural Language Processing (RANLP), 21-23 Sep., 2005, Borovets, Bulgaria*.
- I. Yamada, T. Baldwin, H. Sumiyoshi, M. Shibata, and Y. Yagi. 2007. Automatic acquisition of qualia structure from corpus data. *IEICE - Transactions on Information and Systems archive*, E90-D(10 (October 2007)):1534–1541.