

# Senses of Polysemous Nouns: Building a Computational Lexicon of Basic Japanese Nouns

Wakako KUWAHATA, Minako HASIMOTO  
Information-technology Promotion Agency (IPA)  
3-1-38 Shibakoen, Minato-ku, Tokyo 105, JAPAN  
kuwahata@stc.ipa.go.jp, hasimoto@stc.ipa.go.jp  
Fumihiro AOYAMA  
Obirin University  
3758 Tokiwa-chō Machida-shi, Tokyo 194-02, JAPAN

## Abstract

We have constructed the *IPA Lexicon of Basic Japanese Nouns (IPAL-BN)*, which has a hierarchical structure based on the syntactic and semantic properties of nouns. In our lexicon, each lexical entry consists of *subentries*, and subentries have *semantic property information*. Among these elements, we focus here on the subentry description. Conventional Japanese dictionaries only enumerate various usages. But it is also important to clarify the semantic relations between subentries. Thus we have developed a method for specifying the kind of relationship between subentries, using special cognitive devices such as metaphor, metonymy, and synecdoche. After a brief review of the structure of our lexicon, we discuss how the method can be applied to the lexical description.

## 1 Introduction

The Information-technology Promotion Agency (IPA)<sup>1</sup> has compiled the *IPA Lexicon of the Japanese Language for Computers, Basic Japanese Verbs (IPAL-BV)* (1987) and *Basic Japanese Adjectives (IPAL-BA)* (1990). The *IPAL-BV* contains 861 verbs and the *IPAL-BA* contains 136 adjectives as lexical entries. These lexicons are available for public use and have been widely used in various university and research institute projects that have yielded encouraging results. We started work on the *IPAL-BN* project in 1990. In May 1996, we released the third edition of the *IPAL-BN*, with 1,081 nouns as lexical entries, for the public on networks with FTP service.

The IPAL project is characterized by its linguistic basis. For example, the hierarchical structure

<sup>1</sup>A special juridical body under the jurisdiction of the Ministry of International Trade and Industry, Japan.

of the *IPAL-BN*, which consists of lexical entries, *subentries*, and *semantic property information*, reflects our linguistic considerations concerning the syntactic and semantic properties of nouns. Another example of benefits from our linguistically inspired approach is the description of the kind of *relationship* between subentries. Such information would be useful in various applications, but is not yet explicitly provided in existing Japanese dictionaries. In the following sections, we first briefly introduce the general structure of the *IPAL-BN*, and then describe our method for specifying the kind of relationship between subentries. In the concluding remarks, we also touch on implications of the method for the application systems.

## 2 Structure of IPAL-BN

Figure 1 shows the top-level structure of the *IPAL-BN*. The *IPAL-BN* consists of 1,081 lexical entries.

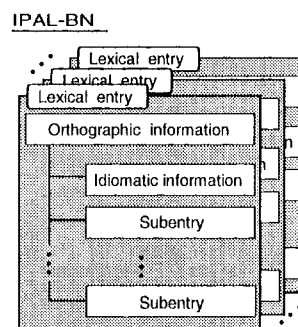


Figure 1: Basic Structure of IPAL-BN

Each lexical entry is composed of *orthographic information*, *idiomatic information*, and *subentries*. The *idiomatic information* accommodates the idiomatic or proverbial uses of the noun that have to be treated separately. Let us take an example, *Hanako-wa hana-ga takai*. This Japanese sentence has the idiomatic meaning, “Hanako is proud,” in addition to the ordinary meaning, “Hanako has a long nose.” In the idiomatic cases, the meaning of the whole sentence cannot be decomposed into the meaning of each word. Thus we reserve the idiomatic information separately from

ordinary meaning sections.

Then we introduce a hierarchy, subdividing each entry that has more than one usage of the word. Each usage is called a *subentry*. The subdivision to subentry is based not only on semantic but also on syntactic characteristics. This categorization process can be illustrated with an example of *hankyō* ‘echo’:

*hankyō* ‘echo’

01. a sound that is reflected off a surface such as the wall of a building.  
*Bōru-no oto-ga heya-no naka-de hankyō-suru.*  
(The sound of the ball echoes in the room.)
02. one’s expression of an opinion about or attitude toward something.  
*Kono hanashi-wa hitobito-no kokoro-ni hankyō-o yobiokosu darō.*  
(This story will arouse an echo in every man’s heart.)

One may note that *hankyō*-01 has a usage in which a noun becomes a verb when followed by “-*suru*”, while *hankyō*-02 does not. On the basis of this difference we divide this noun into these two subentries.

Figure 2 gives an overview of each subentry. A subentry consists of *subentry information* and several pieces of *semantic property information*. The subentry information contains syntactic, semantic, and morphological information common to all parts of the subentry (each semantic property information section). The semantic property information includes syntactic and semantic information. In the case of *hankyō*, we use the syntactic information in the subentry information to describe the difference in the usages (Kuwahata, 1995).

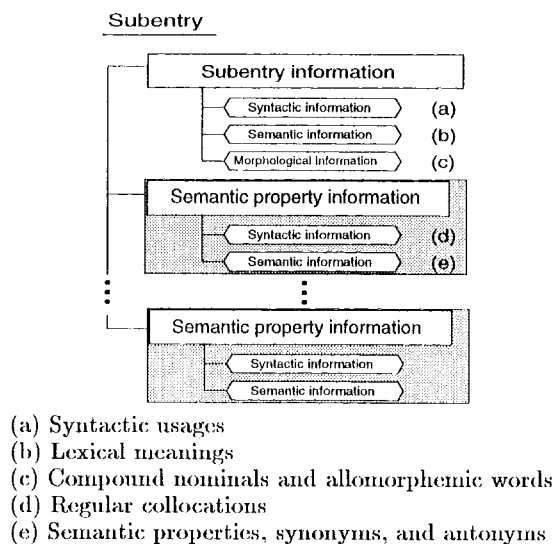


Figure 2: Structure of Subentry

In addition, we examine the subentries in more detail and introduce the concept of the aspects of nouns. For example, ‘the letter’ in ‘I read the letter’ focuses on the *information* in the letter,

whereas its counterpart in ‘I burned the letter’ focuses on the *thing* (i.e., piece of paper) bearing that information. Since we can say ‘I burned the letter that I had read’, the word ‘letter’ does not have two meanings but rather has two aspects. Thus a noun is considered here to have various aspects depending on the predicates used in the sentence containing the noun. These aspects are called *semantic properties* (Aoyama, 1995). Instead of subdividing the lexical entry into multiple subentries, we categorized the regular collocations in each subentry in terms of semantic properties.

For example, let us take the word *ha* ‘tooth/teeth’ which has three semantic properties. In our notation, the semantic properties are labeled by three letters in square brackets. The phrases *ha-o migaku* ‘brush one’s teeth’ and *ha-o nuku* ‘pull one’s tooth’ refer to tooth as a *concrete object* [CON]. And the phrases *ha-ga haeru* ‘cut teeth’ and *ha-ga nukeru* ‘lose teeth’ imply (natural) *phenomena* [PHE], while the phrases *ha-ga jōbu-da* ‘have sound teeth’ and *ha-ga guragurasuru* ‘a tooth feels loose’ single out a condition of teeth from their *potential* conditions [POT].

### 3 Relationships of Subentries

#### 3.1 Problem

Most existing Japanese dictionaries merely enumerate various usages. But clarifying the semantic relations between those usages is important. For example, the noun *tamago* ‘egg/spawn’ has three senses:

*tamago* ‘egg/spawn’

01. an object covered with a hard shell or a membrane, produced by a female animal.  
*Kingyo-ga tamago-o unda.*  
(The goldfish spawned.)
02. a hen’s egg (i.e., some kind of food).  
*Hanako-wa tamago-o ichi pakku katta.*  
(Hanako bought 1 dozen eggs.)
03. a person at the beginning of his/her career.  
*Hanako-wa isha-no tamago-da.*  
(Hanako is a doctor in the making.)

A hen’s egg (02) is one type of object that is covered with a hard shell (01), and (03) is a metaphor with respect to the relation between hen’s egg (02) and hen. Our problem was to explicitly describe these kinds of relations between meanings.

#### 3.2 Approach

In the literature, several attempts have already been made to analyze such semantic relationships<sup>2</sup>. Yamanashi (1995), among others, points out that appreciating such special cognitive devices as metaphor and metonymy is the

<sup>2</sup>Ullmann, 1969; Lakoff and Johnson, 1980; Kuwihiro, 1982; Yamanashi, 1995.

key to understanding polysemy. We employ this method for specifying the kind of relationship between subentries. Currently we note three types of relations:

- **Metaphor**, based on similarities
- **Metonymy**, based on various contiguities
- **Synecdoche**, based on the relation between member and category

**Metaphor** is the similarity-based instrument for extending the meaning of words. At the outset metaphorical expressions are temporarily used figures of speech. However, some metaphors come to be fixed and pass into everyday use. Let us take some examples: *hon-no mushi* ‘a worm of books’ (a person who is crazy about reading), *arasoino tane* ‘a seed of argument’ (a cause for argument), and *kotoba-no kabe* ‘a wall of words’ (a language barrier). We regard these expressions as full-fledged usages of the nouns, and accordingly describe them in the *IPAL-BN*, while temporal usages are not considered for description in our lexicon.

**Metonymy** is an instrument for employing a word to refer to something that distinct from, but is associated in some way with, the original referent of the word. Typical examples are *Nabe-ga oishi*, “The dish is nice,” and *Ano kyatchā-wa i kata-o shiteiru*, “The catcher has nice shoulders (The catcher has a powerful throw).” In the former, the reference has shifted from the container to the content, and in the latter, the reference has shifted from the part of the body to its function.

**Synecdoche** is the instrument that takes the name of a category to stand for one of its members or taking the name of one member to stand for the whole category, as shown in the above hen’s egg example: a hen’s egg is one kind of object that is covered with a hard shell.

It is important not to confuse the connoted relation of synecdoche with the contiguity relation [part]-[whole] of metonymy. Let us take an example of the relation [part]-[whole] to show the difference:

*te* ‘arm/hand’

01. upper limbs of a human (the entirety, consisting of arm and hand).  
“*Hanako-wa ryōhō-no te-o hirogeta.*”  
(Hanako spread her arms.)
02. the end of a person’s arm (the part consisting only of the hand).  
“*Hanako-wa te-o tataita.*”  
(Hanako clapped her hands.)

We do not consider the [part]-[whole] relation between the arm (01) and the hand (02) to be an instance of synecdoche. Let us compare this example of *te* ‘arm/hand’ with the example of *tamago* ‘egg.’ Besides a hen’s egg, there are many other sorts of eggs, such as a turtle’s egg, a pigeon’s egg, and a swallow’s egg. By contrast of

the arm as a whole, only the hand receives the designation *te*. For example, we do not also express the elbow as *te* ‘arm/hand.’ Hence we consider the contiguity relation [part]-[whole] to be one type of metonymy. We distinguish it from the member-category relation of synecdoche<sup>3</sup>.

### 3.3 Method

In this section, we will show some examples of description.

Some metaphorical expressions need some modifier and others do not.

*kata* ‘shoulder’

01. *Hanako-wa kata-o sukumeta.*  
(Hanako shrugged her shoulders.)
02. *Sono uwagi-no kata-wa ōki.*  
(That jacket has wide shoulders.)
03. *Sono yama-no kata-ni yamagoya-ga aru.*  
(There is a lodge on that mountain shoulder.)

(01) is a non-metaphorical use. (02) and (03) are metaphorical uses. One may notice that (02) can be used without *yōfuku-no* ‘of clothes’, but (03) requires *yama-no* ‘mountain-.’ We have the impression that (02) is conventionalized.

We classify metaphorical usages into two types: conventionalized and novel. To the first group we assign expressions that can be used without being modified. Such an expression is indicated by “A => B.” The second group is indicated by “A ...> B.” These mean that A is original and B is metaphorical. We refer to the above relation, hence, as: “01 => 02. 01 ...> 03.”

It is hard to judge which is metaphor if both senses are conventional. Consider:

*shiwa* ‘wrinkle’

01. *Hanako-wa me-no mawari-ni shiwa-ga dekitekita.*  
(Hanako has got wrinkles round her eyes.)
02. *Hanako-wa shatsu-no shiwa-o airon-de nobashita.*  
(Hanako ironed out the wrinkles in her shirt.)

We regard (01) as non-metaphorical and (02) as metaphorical first group.

Various contiguity relations cause metonymical extension. We single out the relation [part]-[whole] as one group, and assign the other contiguity relations to a second group. The first group is indicated by “A >> B” or “A << B” and the second group is indicated by “A [X] -> [Y] B.” “A >> B” means extension from [whole] to [part] and “A << B” means extension in the opposite direction. “A [X] -> [Y] B” means that the relation [X]-[Y] links A to B. These relations include [container]-[content], [material]-[product], [means]-[act].

Here is an example.

<sup>3</sup>Sato (1992) also distinguishes between synecdoche and metonymy. But, Ullmann (1969) and Yamanashi(1995) include synecdoche in metonymy.

*ashi* 'leg/foot'

01. a limb of animal, which includes the foot and is used to support the body and for walking.  
*Ningen-niwa ashi-ga nihon aru.*  
(Humans have two legs.)
02. a part of the foot.  
*Kare-no ashi-wa ōki.*  
(He has big feet.)
03. the way one walks.  
*Hanako-wa ashidori-ga karui.*  
(Hanako has a light feet.)
04. a transport service.  
*Taifu-de ashi-ga ubawareta.*  
(People were deprived of transport by a typhoon.)
05. the leg-like part of a nonliving thing, used to support its body.  
*Sono isu-niwa ashi-ga yonhon aru.*  
(That chair has four legs.)

While (02) and (03) are metonymies with regard to (01), (04) and (05) are metaphors with respect to (03) or (01). We represent these relations as follows: "01 >> 02. 01 [part of a body] --> [function] 03 ...> 04. 01 ...> 05."

Let us now turn to synecdoche. It is indicated by "A  $\supset$  B" or "A  $\subset$  B." The former means that A contains B, and the latter means that B contains A. As mentioned above *tamago* 'egg/spawn' has the following three subentries: (01) an object covered with hard shell or a membrane, produced by a female animal, (02) a hen's egg, (i.e., some kind of food), (03) a person at the beginning of his/her career. While (02) is a synecdoche with regard to (01), (03) is a metaphor with respect to (02). We refer to these relations numerically, thus: "01  $\supset$  02 ...> 03."

## 4 Conclusion

We have compiled the *IPAL Basic Japanese Nouns (IPAL-BN)*, which contains detailed descriptions of syntactic, semantic, morphological, and idiomatic information about nouns by employing a structure that consists of *subentries* and *semantic properties*. Specifically, explicit description semantic relations between subentries is a novel approach in the compilation of Japanese dictionaries for computers.

Nouns are often considered to be simple. But since nouns have various meanings and usages, in Japanese processing systems not all nouns can be handled in a straight-forward way. We believe that the complicated functions of nouns require processing systems that use not only syntactic grammar, which is already employed in conventional natural language processing systems, but also syntactic and semantic information, of the kind contained in the *IPAL-BN* lexicon.

In this paper, we focused on the relationship between subentries and left other relationships untouched. Among the other relationships are idiomatic expressions and semantic properties. For

example, the idiomatic expression *ushi-ga hayai* 'be quick on one's feet' means 'spoil quickly.' We do not mention in the *IPAL-BN* that this example is metaphoric expression, since we include it in idiomatic information outside subentry. To take another example in the semantic property level, the phrase 'contact the police' refers to 'police' as a *organization* [ORG], while in the phrase 'go to the police' it is a *location* [LOC]. The word does not have two meanings but rather has two semantic properties. We also do not describe the relationship between them as one kind of metonymy. Including these points, we expect further consideration is needed to fully describe the senses of polysemous nouns in the lexicon.

## Acknowledgments

The *IPAL-BN* is the result of the efforts of all of our project members, and we are deeply grateful to them. This paper owes much to the thoughtful and helpful suggestions of Akira Honda. Takashi Suzuki and Norihiro Ogata also gave us helpful comments. We also thank Kunio Kashino for his constructive comments on earlier drafts of this paper.

## References

- Aoyama, F., *Sosei-ni motozuku meishi kijutsu-no tame-no wakuqumi. Proc. of IPAL Symposium '95*, pp. 1-9, (in Japanese), IPA, 1995.
- Hasimoto, M., Kuwahata, W., Murata, K., Aoyama, F., and Tonoike, T., *Some remarks on ways to compile Japanese lexicons for computers, Proc. of the International Workshop on Sharable Natural Language Resources*, pp. 115-122, 1994.
- Kunihiro, T., *Imiron-no hōhō* (in Japanese), Taishukan-shoten, 1982.
- Kuwahata, W., Hasimoto, M. and Murata, K., *Construction of IPAL-BN:IPA Lexicon of Basic Japanese Nouns, Proc. of NLP'95 vol. 1*, pp. 137-142, 1995.
- Lakoff, G. and Johnson, M., *Metaphors We Live By*, The University of Chicago, 1980.
- Sato, N., *Retorikku-kankaku* (in Japanese), Kdansha Gakujutsu Bunko Ban, 1992.
- Ullmann, S., *Semantics: An Introduction to the Science of Meaning*, (Japanese translation by Y. Ikegami, published by Taishukan-shoten, 1969).
- Yamanashi, M., *Ninchi bunpou ron* (in Japanese), Hituzi Syobo, 1995.

## Appendix: How to get IPALs

The *IPAL-BN* and other *IPALs* are available for public use. Anyone interested in using the *IPALs* is invited to contact one of the authors of this paper. Residents outside Japan may obtain the copies of the manuals (written in Japanese) for the *IPALs* free of charge.