RETHINKING METALINGUISTIC LABELS: SPATIO-TEMPORAL METALINGUISTIC TERMS IN LEARNERS’ DICTIONARIES OF JAPANESE EXPRESSIONS

Andrej BEKEŠ
University of Ljubljana
andrej.bekes@gmail.com

Abstract
Paper examines spatio-temporal metalinguistic terms in learners’ dictionaries of Japanese expressions in major existing dictionaries. Based on the analysis it proposes a layered metalinguistic labelling solution to achieve the greatest efficiency with the smallest possible number of labels being employed.

Keywords: Japanese; metalinguistic label; semantic/functional index; grammar/ expression dictionaries

1. Introduction

Efficient semantic labelling is essential if a dictionary is to be used for the composition as well as the reception of texts. When creating a learners’ dictionary, this task tends to be done based on experience and tradition, and often in quite an ad hoc fashion, that is without systematic theoretical considerations (e.g. Group JAMASSY 1998 where I was in charge of semantic/functional labels).

One option for a more systematic treatment is offered by the NSM (natural semantic metalanguage) approach (e.g. Goddard and Wierzbicka 2007, for a critical assessment c.f. Trobevšek Drobnak 2009). Another possibility is to build an array of metalinguistic labels from the bottom up, from corpora (e.g. Labrador De La Cruz 2004). On the other hand, the chosen labels should also be as theoretically relevant as possible.
This means that if adequate labels are to be produced, not only collocations but also broader semantic properties, implicit within the wider context, need to be considered.

Languaging (e.g. Becker 1988) is an activity where our linguistic potential is applied to fluid on-going situations. Not only functional expressions but also lexical items such as verbs, adjectives and nouns, often have multiple meanings, which arise through interplay with the context. The problem is how to capture these characteristics in a metalanguage intended for explanation of the lexicon, or when providing learners with semantic/functional labels.

One of the drawbacks of NSM is its apparently static view of the lexicon and the meanings of lexical items, missing a crucial property of human communication. Regarding methodology, the intuitively chosen approach of Group JAMASSY (1998) to provide meanings not so much by description as by illustration, giving many examples of particular uses of functional lexical items, seems appropriate. While more terse than Makino and Tsutsui (2008) in the sense that its explicit explanations are limited to short commentaries, it seems to be a less restrictive approach (than what?) when guiding experienced users(of what?)/learners.

The context-based approach, resonating with the basic argument in Labrador De La Cruz (2004), has recently been taken up systematically. Limiting the discussion just to the field of Japanese language learning support tools, two outstanding such tools should be noted. One is Natsume (http://hinoki.ryu.titech.ac.jp/natsume/), developed by Kikuko Nishina and associates and running partly on the BCCWJ and partly on closed corpora, and the other is NINJAL LWP (Lago Word Profiler), developed by Prashant Pardeshi and associates and implemented using the BCCWJ and TWC corpora (http://corpus.tsukuba.ac.jp/search/). These tools provide users with usage based information, extracted online by means of an analysis of the chosen corpora. For this reason, these tools are actually used not only by intermediate and advanced learners, but also by professional dictionary makers.

In spite of the power of these new tools, well thought-out dictionaries for beginner and lower intermediate learners are also crucial to speed up efficient learning of Japanese. For this purpose, an efficient array of semantic/functional labels that guides the users towards an effective understanding of the lexical item is necessary. The labels should be chosen so that for composition, they can efficiently guide the learner towards those expressions in the target language that are most appropriate to achieve the learner’s goal in a given situation. Since learners are not translators, however, this goal should be achieved through learners’ intuitive grasp “from within” of how the target language works.

This paper focuses on spatio-temporal metalinguistic terms, with an emphasis on methodological issues, and analyses the implementation of such metalinguistic terms in contemporary Japanese grammar/expression dictionaries, in particular Group JAMASSY (1998). Building on the analysis of how spatio-temporal terms are structured
in Group JAMASSY (1998) I argue for a concise but systematic and pragmatically relevant approach in semantic labelling.

2. Japanese grammar / expression dictionaries

A number of dictionaries (Group JAMASSY, Makino and Tsutsui, Tomomatsu et al. etc.) have been published since 1990 with the express purpose of easing the learning of items that do not appear in traditional dictionaries, i.e., functional expressions, sentence patterns, etc. In this paper, two such dictionaries were chosen, based on their breadth of coverage and popularity. An additional reason is that I am a co-author of one (Group JAMASSY 1998) and this paper serves as a re-examination of that previous work.


The purpose of the dictionaries is largely similar: providing clear information to learners about common expressions and patterns.

Makino and Tsutsui authored three volumes, from elementary to advanced, while Group JAMASSY’s dictionary targets upper intermediate and advanced level learners. Therefore, only Makino and Tsutsui’s third volume (2008), targeting advanced level learners, will be taken into consideration here.

In both dictionaries the main entries include lexical items and constructions, and are organised alphabetically (Makino and Tsutsui) or in gojūon order (Group JAMASSY).

Makino and Tsutsui offer detailed information on formation, i.e., the constructions in which particular expressions are typically used. Group JAMASSY handles this type of information in subentries that are often detailed and hence sometimes difficult to follow.

There is one crucial difference in the organisation of the dictionaries. Makino and Tsutsui is bilingual - with Japanese entries and examples but strictly English explanations and glosses. On the other hand, Group JAMASSY’s dictionary was conceived as a monolingual dictionary, aimed primarily at upper intermediate and advanced learners and as a reference book for Japanese language teachers. In the wake of its success, translations of the dictionary into Chinese, Korean, Thai and Vietnamese have appeared.

While Makino and Tsutsui provide ample examples of use together with translations and detailed explanations, Group JAMASSY relies on the power of multiple examples, supplemented with direct, concise explanations. In the translated versions of the dictionary the examples and the explanations are translated into the respective language, resulting in a dictionary of a structure similar to Makino and Tsutsui.
Another difference is in the organisation of the index. Makino and Tsutsui (2008), being bilingual, offers two indexes, an index of romanised Japanese entries in alphabetical order, and an index of English translations of entries, also in alphabetical order. This solution may seem helpful for users engaging in productive tasks, but is not always conducive to a quick location of the desired expression in Japanese.

Group JAMASSY tries to cope with the problem of quick entry location by including two indexes: besides the gojūon index of entries, there is also an index of entries sorted by their meaning and/or function. This semantic/functional index consists of more than one hundred metalinguistic labels, referring broadly to the meaning and/or function of a particular entry or subentry. In this respect, Group JAMASSY (1998) is similar to some of the other grammar dictionaries that appeared later, such as Tomomatsu et al. (2007/2010).

Below, this paper focuses on the subset of these labels that refer to spatial concepts, temporal concepts, or both, in Group JAMASSY (1998).

3. Metalinguistic labels

Attempts at creating a language that would precisely convey meaning across languages have a distinguished lineage, going back to the tradition of searching for a universal language - with G. W. Leibnitz to be mentioned as one prominent proponent of the idea (e.g. Yaguello 1990, Eco 2003), and with NSM (e.g. Goddard and Wierzbicka 2007) being a notable present attempt.

In both dictionaries under scrutiny, the explanations should have ideally been based on some subset of language in the vein of NSM, but this is not the case with Group JAMASSY and does not seem to be the case with Makino and Tsutsui either. In both dictionaries ordinary, though to some extent controlled, language is used for the explanations.

3.2 The purpose of the metalinguistic labels in Group JAMASSY dictionary

Group JAMASSY (1998) uses metalinguistic labels to hint at the meaning and/or function of entries and subentries. There are two types of labels: (a) labels used in the semantic/functional index, to direct the learner towards the relevant entries during language production tasks, and (b) those used in the main body of the dictionary, to disambiguate different meanings and usages.

The problem is that both sets of labels are not completely identical and are used in a way that seems somewhat inconsistent. Let us consider for example the entry あいちが.

(1) a

期間
Rethinking Metalinguistic Labels: … 13

あいだ ........... p. 2
あいだに ........... p. 2
空間的関係
のあいだ ........... p. 2

b

【あいだ】
1 N のあいだ
a N のあいだ ＜空間＞
........
2 あいだ ＜時間＞
a ...あいだ
[N のあいだ]
[A-いあいだ]
[V-ている/V-るあいだ]

In the semantic/functional index the entry あいだ is listed under the label ＜期間＞ (meaning interval of time, period, see (1)a above), while in the main body, あいだ is disambiguated with the synonymous / hypernymous ＜時間＞.

In the index one also finds labels such as ＜無関係＞, ＜前後関係＞ and ＜空間的関係＞, while in the body of the dictionary, as in example in (1)b, the hypernym ＜関係＞ is used, which in this context refers to the specific meaning of a “human relationship” and happens to be the most frequent use in TWC.

Also, while for example ＜空間＞ and ＜時間＞, as above, are often used in the main body of the dictionary, there are no such labels in the index. Here, on the other hand, one can find ＜時点＞, ＜期間＞, ＜空間的関係＞, which point to the relevant entries in the body.

In this way, there are 158 labels used only in the main body, 115 labels used only in the index, and 45 labels that are used both in the index and in the main body of the dictionary.

3.3 Clashing requirements

There is a contradiction inherent in devising metalinguistic expressions to be used for explanation in dictionaries. On the one hand, the number of such expressions has to be small, so that they are manageable, and on the other hand, for these expressions to be specific enough, their number has to be sufficiently large. This is the clash of requirements that Leibnitz and others had to face early in their work on universal language (e.g. Eco 2003; Yaguello 1990). The problem that for a while also hindered attempts at automatic translation is that if one wants to make a language with the expressive power of natural languages it has to be as complex as natural languages.
Group JAMASSY’s attempt was no exception. Its goal was simplicity and intuitive ease of use, which resulted in plentiful examples, concise explanations, and a very limited set of non-technical transparent semantic/functional labels used in the main body and index of the dictionary. The request that labels in the index should be specific enough to direct the learner to a small set of potential candidate entries is in conflict with the requirement that the number of labels should be kept at a manageably small number. The outcome of the clash was the aforementioned inconsistency between the labels used in the index and those used in the body of the dictionary.

Though this is theoretically an unresolvable problem (NSM being no exception), a trade-off is possible. The compromise would be to establish an array of labels sufficiently rich for practical purposes but at the same time also concise enough to be manageable.

Randomly chosen examples such as (1) above show that an improvement in the consistency and accuracy of semantic/functional labels is possible.

4. Some proposals for improvement

Limiting the scope to spatio-temporal labels, I will illustrate a few possible paths out of this impasse.

4.4 Consistency of use - labels in the body and in the index

There are many examples where the use of semantic/functional labels could be made more consistent by treating the labels in the body of the dictionary and the labels in the index as parts of the same set. In the following sections, examples based on the use of spatio-temporal labels are presented.

4.4.1 Unifying labels in the dictionary body and in the index

The labels in the index and the labels used for disambiguation in the body of the dictionary are sometimes related (hypernym-hyponym etc...), yet differ. For example, there are cases when 期間 is used in the index while 時間 is used in the body of the dictionary. The first entry in the dictionary is a good example:

(2) 2 あいだ <時間>  p. 2
a ...あいだ
[N のあいだ]
[A-いあいだ]
[V-ている/V-るあいだ]
b ...あいだに
[N のあいだに]
Rethinking Metalinguistic Labels: … 15

[Na なあだに]  
[A-いあだに]  
[V-ている/v-るあだに]

In (2) 2a, the meaning of あだ can be glossed as "an interval of time during which an action continues" (継続動作が続く/いた期間). It would seem to be more user friendly to use the same label for disambiguation and in the index, and 期間 is more specific than 时间. In 2b, the meaning of あだに can be glossed as "(relatively) instantaneous action, taking place within an interval of time" (瞬間動作が起る/った期間), so again, as in 2a, it seems appropriate for the labels to be the same, i.e., 期間. A more or less identical situation seems to be the case in (3) below, semantically similar to the preceding example, with the meaning of うちに glossed as "(relatively) instantaneous action, taking place within an interval of time" (瞬間動作が起る/った期間):

(3) 2 うち <時間>  
[Nのうちに]  
[Na なうちに]  
[A-いううちに]

In example (4) and (5) below, the situation is similar again. The meaning of both constructions can be glossed as "an interval of time during which an action or state continues" 状態・継続動作が続く/いた期間. In (4), as in (2)2a, using 期間 for disambiguation would be more appropriate.

(4) 2 N ごしの <時間>  
(5) 2 N じゅう <時間>

On the other hand, in (6), both keeping and changing the disambiguation label could be argued for:

(6) 3...まで <時間>  
a N まで  
b V-るまで

The gloss of まで in 3a in (6) is “the time until which a state or some activity continues” (状態・継続動作の時限), and the gloss of まで in 3b is “the time limit defined by an action, until which a state or some activity continues” (態・継続動作の、別の動作によって限定される時限). In both cases, the hypernym 時間 seems acceptable from the point of view of label economy, though a more specific label, such as “time limit” 時限 would be even more appropriate. Additionally, in 3b, the temporal-relational semantic relationship between the defining action and continuing state or activity could also be emphasized, by means of double labelling, a possibility discussed in more detail in Section 3.2.
Similarly, the construction in the example below:

(7) 2 N を N にひかえて <場所> p. 505

expresses the general meaning of a spatial relationship between a location and its surroundings. Therefore the more specific “place, location” <場所> could be replaced by <空間> as a disambiguator.

4.4.2 Consistent disambiguation of subentries

Consistent disambiguation of the subentries would make searching easier and improve understanding of the entries. The first entry in the dictionary is once again a good example.

(8) 【あいだ】
   1 N のあいだ p. 2
   a N のあいだ <空間>
   b N のあいだ <関係>
   2 あいだ <時間>
   a ......あいだ
   b ...あいだに
   ...

   In (8) the macrostructure of the entry 【あいだ】 is shown. While the temporal use of あいだ is disambiguated at the first level of the subentry, the spatial and relational use are disambiguated at the second level of the subentries. Restructuring the entry 【あいだ】 into

(8’) 【あいだ】
   1 N のあいだ <空間>
   2 N のあいだ <関係>
   3 あいだ <時間>
   a .....あいだ
   b ...あいだに
   ...

would render the treatment of subentries in the dictionary more consistent.

Restructuring the entry would also be advisable in cases such as 【まで】 below:

(9) 【まで】 p. 545
   1 N から N まで <∅>
   (1) シンポジウムは1時から3時まで第3会場で行います。
   (2) A：大阪から東京までどのくらいかかりますか。
In (9) the first level subentry 1 is not disambiguated at all, while giving examples of the temporal and spatial use of から まで. Such situations may discourage an inexperienced user, who otherwise could make good use of such a dictionary, if its entries were organised in a more consistent and transparent way. A more appropriate entry structure would be:

(9')

【まで】

1 N から N まで <時間>
(1) シンポジウムは 1 時から 3 時まで第 3 会場で行います。

2 N から N まで <空間>
(2) A: 大阪から東京までどのくらいかかりますか。...

3 N まで <目的地>
4...まで <時間>
a N まで
b V-るまで
5...まで <程度>
...

4.5 Combination of labels for greater precision

One way to increase the expressive power of a limited number of semantic/functional labels is to combine the labels. For example, a superordinate label could specify a wider semantic area, while the subordinate label of the same entry would narrow the focus to one of its aspects. Incongruent examples, using a partly overlapping set of labels in the index and in the main body, such as (10) below, provide a hint:

(10) index: 条件(確定条件); 条件(仮定条件); 条件(一般条件)
【するとと】
【とすると】
主要 body:
1 ... とすると <仮定条件>
2 ... とすると <確定条件>

The index entry, which at present looks like:

(11) 条件(一般条件) と
would look like:

(11’) 条件
  <一般条件> と
  ...
  <仮定条件> かりに /...たら/...ば
  ...
  <確定条件> とすると

while the labelling in the main body of the dictionary would not change. With a two-layered labelling in the index the user would have a clearer idea of how entries are interrelated.

Below I offer some additional examples of temporal and spatial labels that could become more transparent by being rearranged into two layers.

In the index entry (12) below, the spatial use of the entry expression ごし is indicated by the label 空間的関係, while there is no label provided to indicate the temporal use of ごし.

(12) 【ごし】 p. 110
  1 N ごし <空間>
  2 N ごし <時間>

In the “spatial” use of ごし the relational meaning seems to be more general, and it should take precedence over the spatial meaning. As for the temporal use of ごし, as shown in example (4) above, it implies an interval. The temporal meaning should therefore be labelled as 期間. The index should hence be rewritten as follows:

(12’) 関係
  <空間的関係>
  ごし
  ...
  期間
  ごし
  ...


The main entry would then look like:

(12’’)  Curtis 【ごし】

1 N ごし <空間的関係> → <N(物) を間において>の行為
2 N ごし <期間> → 状態・継続動作が続く/いた期間

There is no end to possible examples. The next example is again an entry with both spatial and temporal meanings.

(13)  Curtis 【じゅう】 p. 145

1 N じゅう (に) <空間>
2 N じゅう <時間>

In the index, the entry じゅう is referred to by the labels 空間的関係 and 期間. In its spatial use, the meaning seems not to be so much relational as implying “the range or scope of some activity” (動作が及ぶ範囲、空間), thus 範囲 is appropriate, subspecified for space <空間>. The temporal use label, <時間>, seems too vague for the gloss “the period while a situation or an action continues” (状態・継続動作が続く/いた期間). Both the spatial and temporal uses are related; 期間 is a sort of temporal range, that is, a 範囲. The revised index labelling and entry should thus look as follows:

(13’)  index:

範囲

<空間>

じゅう;

期間

じゅう

entry:

 Curtis 【じゅう】

1 N じゅう (に) <範囲-空間>
2 N じゅう <期間> → 状態・継続動作が続く/いた期間

In this section, a two-layered labelling was proposed. As can be seen from example (14), finer subdivisions with more layers of labels are possible and sometimes necessary. In (14) the entry 【あと】 is referred to in the index by two labels, 空間的関係 and 前後関係. The problem is that 前後関係 can refer to both temporal and spatial order, so as a label it is not very useful for the learner.

(14)  Curtis 【あと】 p. 8, 9

1 あと <空間>

[N のあと]

[V-る/ナ-たあと]

2 あと (で/に) <時間>

a ...あと

[N のあと] 動作(N)の継起
Two-layered labelling would provide clearer information for the learner. The entry あと 1 is about temporal and spatial order, so rather than 時間 or 空間, its relational property should be given priority in the choice of labels. In the index, the labels should be arranged as below, in (14’). For greater precision and a sharper focus, an additional layer can be added, in this case, <前後関係>. This is the reverse of the original index, where the ambiguous label <前後関係> appears as the first label. Thus we would have in the index:

(14’) 

関係
   <空間的関係>
   <前後関係>
あと
   <時間的関係>
   <前後関係>
あと

and in the body:

(14’’) 

【あと 1】
1 あと <空間的関係-前後関係>
   [N のあと]
   [V-る/V-たあと]
2 あと(で/に) <時間的関係-前後関係>
   a ...あと
   [N のあと]
   [V-たあと]

4.6 Contribution of corpora

Frequency data from corpora can be a guide for ordering different usages, though this principle is not an absolute guide. Let us consider entry (1) once again in (15).

(15) 

【あいだ】
1 N のあいだ
   a N のあいだ <空間>
   ...........
2 あいだ <関係>
   b N のあいだ <時間>
   a ...あいだ
   ...........
Here for example frequency based ordering and semantic (more general first) ordering clash. Data from Lago TWC show that the use as RELATION is more frequent than the use as SPACE. On the other hand, semantically, SPACE is primary and RELATION is derived.

Further, the semantic properties of N in N のあいだ may also be relevant. The choices in the Group JAMASSY do not always seem to be the most appropriate. Here are some representative examples, based on the same entry:

(16) b N のあいだ〈関係〉 p.2 <Group JAMASSY>
(1) 最近二人の間はうまくいっていないようだ。
(2) そのホテルは安くて清潔なので、旅行者たちの間で人気がある。
(3) 二つの事件の間にはなにか関係があるらしい。

Considering the first example above, (16)(1), frequencies in LAGO TWC are 402 for N=二人 and 1535 for N=人々. This difference should be big enough to warrant the choice of 人々 instead as the typical example of relation involving members of a set. On the other hand, the second example of use (16)(2) seems to be appropriate as far as the choice of N is concerned, as Ns that belong to the semantic category HUMAN (<人間>) represent the vast majority of examples of use in LAGO TWC. There is doubt, however, surrounding whether (16)(2) is indeed an example of RELATION <関係>. It seems that could be interpreted rather as expressing a RANGE (or scope) <範囲> where the predication relationship is between 「ホテル」 and 「人気がある」. The third example of use, (16)(3) shows a relation between abstract Ns. The frequency in LAGO TWC for N= <abstract noun> is very small (< 100), compared to N=人々 (1535), but the example seems to be appropriate, because abstract nouns are a separate category.

In accordance with Labrador De La Cruz (2004), labels based on corpora may improve the overall quality of labelling. Abstractions should be based on empirically obtained examples of use and on the goals of a particular dictionary. Some examples of あいだ from LAGO TWC follow:

(17) a 夫婦の間がギクシャク。 <frq. 583>
 b 異なる国の間で行う売買取引。 <frq. 209>
 c この「縁と契約」という二つの文化の間には、天と地ほどの乖離がある。

Here, (17) a is a relation between two specific sentient beings with volition, also quite a frequent pattern observed in the corpus. (17)b is a relation between two possibly not well defined human groupings, but still rather frequent. (17)c is a relation among abstract categories, denoting general characteristics of some human society, and such constructions are not at all frequent in the corpus. From the three examples in (17) it is possible to see that the general label RELATION could be subspecified further as VOLITIONAL (in 17a), as GENERALISED VOLITIONAL (in 17b) and as PERCEIVED/ASCRIBED (in 17c). How far this labelling process should go in practice
is perhaps one of the most difficult tasks of future dictionary making, not only for JAMASSY 2.0.

5. Conclusions

In Section 1 I compared the organisation of Group JAMASSY (1998) with that of Makino and Tsutsui (2008), pointing out the differences stemming from the different basic concepts on which the dictionaries are based. Makino and Tsutsui’s volumes are more like grammar handbooks, arranged in dictionary form, with ample examples and explanations. On the other hand Group JAMASSY’s dictionary is organised as a dictionary of entries that are not described in general dictionaries. While the two dictionaries overlap, Group JAMASSY covers a larger number of entries. It offers plentiful examples with concise explanations. It also offers relatively systematic semantic/functional labels, used in the index to guide the user towards the desired entries, and within the entries, to disambiguate and explain usage.

In section 2, I briefly examined some issues connected with metalinguistic labels, in particular the inherent limitations of “universal” semantic/functional labels, and clashing requirements that force editors to find a pragmatic balance between accuracy and usability.

Section 3 dealt with the metalinguistic labels in the Group JAMASSY dictionary. Consistency of structuring and use is a precondition for the successful application of labels. Some inconsistencies in the organisation and choice of labels were pointed out and alternatives suggested. Further, having two sets of labels, one for the index and one for the main body of the dictionary, was shown to be inappropriate. As a solution, a merger into one consistently organised set was proposed, such that it could be used to create consistent disambiguation of entries and subentries, while using the same labels also for the meaning and function index. The merger of labels has a drawback, i.e., an increase in the number of labels. To keep the number of labels low and at the same time efficient for accessing the entries from the index via meaning/function, or as transparent disambiguators in the main body of the dictionary, a layered use of labels was proposed and illustrated with examples of spatio-temporal entries. Lastly, some possibilities of how to use corpus data were also outlined.

Literature


