COMPILATION OF JAPANESE BASIC VERB USAGE HANDBOOK FOR JFL LEARNERS: A PROJECT REPORT

Prashant PARDESHI
National Institute for Japanese Language and Linguistics (NINJAL)
prashantpardeshi@gmail.com

Shingo IMAI	Tsukuba University
Kazuyuki KIRYU	Mimasaka University
Sangmok LEE	Kyushu University
Shiro AKASEGAWA	Lago Institute of Language
Yasunari IMAMURA	National Institute for Japanese Language and Linguistics (NINJAL)

Abstract

In this article we introduce a collaborative research project entitled "*Nihongogakushuushayou kihondoushi youhouhandbook no sakusei* (Compilation of Japanese Basic Verb Usage Handbook for Japanese as Foreign Language (JFL) Learners)" carried out at the National Institute for Japanese Language and Linguistics (NINJAL) and report on the progress of its research product, namely, a prototype of a basic verb usage handbook (referred to as "handbook" below). The handbook differs in many ways from the conventional printed dictionaries or electronic dictionaries available at present. First, the handbook is compiled online and will be made available on internet for free access. Secondly, the handbook is corpus-based: the contents of the entry are written taking into consideration the actual use of the headword using the BCCWJ corpus. Also, it contains illustrative examples of particular meanings culled from the BCCWJ corpus as well as those coined by the entry-writers. Third, the framework used in the description of semantic issues (polysemy network, cognitive mechanism underlying semantic extensions and semantic relationships among various meanings, etc.) is cognitive grammar, which adopts a prototype approach. Fourth, it includes audio-visual contents (such as audio files and animations/video clips etc.) for effective understanding, acquisition and retention of various meanings of a polysemous verb. Fifth, the handbook is bilingual (Japanese-Chinese, Japanese-Korean and Japanese-Marathi) and incorporates insights of contrastive studies and second language acquisition. The handbook is an attempt to share cutting edge research insights of various branches of linguistics with Japanese language pedagogy. It is hoped that the handbook will prove to be useful for JFL learners as well as Japanese language teachers across the globe.

Keywords

basic verbs; corpus-based; cognitive grammar; audio-visual contents; bilingual dictionary; multilingual dictionary

Izvleček

Članek predstavlja skupinski raziskovalni projekt z naslovom "*Nihongogakushuushayou kihondoushi youhouhandbook no sakusei* (Izdelava priročnika o rabi japonskih osnovnih glagolov za učence japonščine kot tujega jezika)", ki poteka na Državnem inštitutu za japonski jezik in jezikoslovje (National Institute for Japanese Language and Linguistics - NINJAL), ter poroča o trenutnem stanju raziskovalnega izida, t.j. prototipa priročnika o rabi osnovnih glagolov (v nadaljevanju "priročnik"). Priročnik se v marsičem razlikuje od običajnih tiskanih in elektronskih slovarjev, ki so trenutno dosegljivi. Prva značilnost je ta, da se priročnik ureja preko spleta in bo prosto dostopno objavljen na spletu. Druga je ta, da je priročnik osnovan na korpusih: pri redakciji gesel se upošteva dejanska raba iztočnic v korpusu BCCWJ, priročnik pa vsebuje tako primere rabe posameznih podpomenov, ki se črpajo iz korpusa BCCWJ, kot tudi primere, ki jih sestavijo redaktorji. Tretja značilnost je ta, da se semantični vidiki (pomenske mreže, kognitivni mehanizmi, ki botrujejo pomenskim širitvam, ter pomenske povezave med posameznimi podpomeni, ipd.) opisujejo v okviru kognitivne slovnice s prototipnim pristopom. Četrta značilnost je ta, da vključuje zvočne in slikovne vsebine (zvočne posnetke, animacije, videoposnetke ipd.) kot pomoč pri učinkovitem razumevanju, učenju in pomnjenju različnih pomenov večpomenskih glagolov. Peta značilnost je ta, da je priročnik dvojezičen (japonsko-kitajski, japonsko-korejski in japonsko-maratski) in vključuje spoznanja protistavnega jezikoslovja in vede o učenju tujih jezikov. Priročnik je poskus zlitja najnovejših raziskovalnih spoznanj različnih vej jezikoslovja z didaktiko japonskega jezika. Upamo, da bo priročnik koristil tako učencem kot učiteljem japonščine po celem svetu.

Ključne beside

osnovni glagoli; korpusno osnovan; kognitivna slovnica; zvočno-slikovne vsebine; dvojezični slovar; večjezični slovar

1.   Introduction

Verbs as predicators are one of the crucial components determining the skeleton of a sentence, which serves as a basic unit of communication. For improving communication skills in Japanese it is imperative for JFL (Japanese as foreign language) learners to master various usages of basic verbs used frequently in day-to-day communication in a systematic way. At the National Institute for Japanese Language and Linguistics (NINJAL), a collaborative research project entitled "*Nihongogakushuushayou kihondoushi youhouhandbook no sakusei* (Compilation of Japanese Basic Verb Usage Handbook for Japanese as Foreign Language (JFL) Learners)" is being carried out (project leader: Prashant Pardeshi, timeline: October 2009-September 2012). The aim of the project is to develop a prototype for the compilation of a handbook of usage of basic verbs in Japanese frequently used in day-to-day conversation by integrating state-of-the-art insights from various related fields such as Cognitive Linguistics, Corpus Linguistics, Japanese Linguistics, Japanese Language Pedagogy, Contrastive Linguistics, and Linguistic Typology. The envisaged end product is a set of small-scale bi-lingual handbooks such as Japanese-Chinese, Japanese-Korean and Japanese-Marathi, compiled adopting the prototype developed in the project. We believe that such a bilingual handbook of usage of Japanese basic verbs

would be of great help for JFL learners in their effort to acquire the Japanese language systematically and efficiently.

The handbooks under compilation differ from existing dictionaries in various respects such as compilation policy, scope and contents of description and the writing and editing process. In this article we report on the progress of the project and salient features of its envisaged research output, namely, a prototype of a bilingual Japanese basic verb usage handbook (referred to as handbook below). The structure of this article is as follows. In section 2 we provide the outline of the handbook project and a overview of the salient features of the handbook under preparation. Against this backdrop, in section 3 we exemplify the organization of each entry with the help of a concrete example – the verb *hashiru* "to run" – and describe the (tentative) methodology of description. One of the salient features of the handbook is that it is corpus-based. In section 4, we describe the tools/interfaces developed for retrieving information necessary for writing an entry from the corpora of correct use of Japanese and of the errors of JFL learners. Further, the compilation and editing work of the handbook is carried out online using a web-based editing tool. In section 5, we describe the multilingual editing tool developed in this project. This tool allows us to transcend the barriers of space and time. Furthermore, we are developing audio-visual contents in order to foster understanding of various meanings of polysemous verbs. In section 6 we introduce those contents. Finally, in section 7 we discuss future prospects.

2.   Overview of the handbook project and salient features of the handbook

2.1 Overview of the handbook project

We believe that systematic learning of polysemous basic verbs including features such as the semantic behaviour (semantic extensions of a verb and interrelationship among its various meanings, related words such as synonyms, antonyms etc., proverbs/idioms involving the verb in question etc.), grammatical/syntactic behaviour (voice and polarity bias, aspectual and modal characteristics, co-occurrence restrictions, modifiers/adverbial elements, ungrammatical/unnatural usages, etc.), argument structure (case frame), genre/register bias, etc. is necessary in order to master communication skills in Japanese. Further, it is also necessary to know where and how the Japanese language (target language: L2) is similar to or different from the user's mother tongue (source language: L1). In view of this, the aim of the project is to develop a prototype for the compilation of a handbook of usage of Japanese basic verbs by integrating state-of-the-art insights from various related fields of theoretical and applied linguistics for the JFL learner. At present, 58 scholars from various parts of the globe are participating in this project. Out of these 58 scholars, 42 are native speakers

of Japanese while 16 are non-native researchers working on Japanese language for a long period of time[1].

Since the primary goal of the project is qualitative, viz. developing a prototype of a bilingual basic verb usage handbook, we decided to restrict the quantity (number) of verbs and focus on highly polysemous basic verbs which pose a great challenge for JFL learners. Concretely speaking, we focus on the following 11 verbs: verbs of spatial motion (vertical motion: *agaru* "go/move up", *ageru* "cause to go/move up", *sagaru* "go/move down", *sageru* "cause to go/move down", and horizontal motion: *hashiru* "run"), and verbs of temporary or permanent transfer of possession (*ageru* "to give something to someone as a present/gift", *morau* "to receive something from someone as a present/gift", *uru* "to sell", *kau* "to buy", *kasu* "to lend" and *kariru* "to borrow"). All of these verbs are highly polysemous: for example, in our handbook there are 19 meanings/senses for *agaru* "go/move up", 22 for *ageru* "cause to go/move up", and 11 for *hashiru* "run". In section 3, we describe the policy and method of description of an entry through the example of the entry for *hashiru* "run" in our handbook.

2.2 Salient features of the handbook

The handbook under preparation is in electronic online form and the target users of the handbook are envisioned to be advanced JFL learners and native as well as non-native teachers of Japanese. In addition to the dictionary-like usage for looking up the meaning and examples illustrating various meanings of a verb, the handbook serves as a reference grammar also containing many salient features such as explanation of cognitive mechanisms underlying semantics extensions, notes on grammatical and non-grammatical usages, pragmatics or context-related explanations, tips from the contrastive perspective (comparison with the L1 of the JFL learner), "real" examples from the corpus, visual contents such as image-schema (static, abstract line drawings as well as concrete animations and video-clips), and audio-contents such as accent pattern and sound-files for all illustrative examples. Further, the descriptions and "coined" examples are all based on the actual use of the verb as "objectively" gleaned through the corpus data.

Out of all these salient features, two features can be considered as "discriminatory" ones that set apart the present handbook from the bi-lingual dictionaries available at present: (i) corpus-based approach: drawing on a corpus of "correct use" of Japanese native speakers and one of "erroneous use" of JFL learners in addition to the intuitions of scholars for the composition of entries and (ii) incorporating the insights of cognitive linguistics and contrastive linguistics.

For the corpus of "correct use" of Japanese native speakers we used the BCCWJ corpus (Maekawa, 2012) developed by the National Institute for Japanese Language

---

[1] For further details visit the project HP: http://www.ninjal.ac.jp/research/project/b/youhoujiten/.

and Linguistics (NINJAL) and developed an interface called NINJAL-LWP for the BCCWJ corpus (NLB) to cull the information necessary for writing a entry. For "erroneous uses" of JFL learners we used the data from Teramura (1990) and developed a interface to retrieve relevant information from it (see section 4 for details). The prototype of the handbook under preparation incorporates examples from BCCWJ corpus culled with the help of NLB and thus offers both coined as well as real examples side-by-side (see the tentative design in Figure 1).

For incorporating the insights of cognitive linguistics we have incorporated visual contents such as image-schema (static, abstract line drawings as well as concrete animations and video-clips), and audio-contents such as accent pattern and sound-files for all illustrative examples taking full advantage of the web-based nature of the handbook. As for incorporating insights of contrastive linguistics, in addition to grammatical similarities and differences between Japanese and JFL's native language we have provided extra-grammatical information such as notes on pragmatics and cultural factors.

The handbook is compiled/edited using a web-based editing tool connecting scholars in Japan, China and India. Such a handbook differs in many respects from contemporary bilingual dictionaries and therefore we purposely call it a bilingual handbook. In the following sections prominent salient features of the handbook are discussed.

3.   The organization of an entry and the (tentative) methodology of description

3.1 Organization of an entry

The organization of an entry/headword is explained below with the help of the concrete example of the verb *hashiru* (to run). Following this, the methodology of description is mentioned. However it should be borne in mind that the statement pertaining to the methodology of description is tentative and subject to change.

〔アクセント：Accent〕**LHL**

〔活用：Conjugation〕**hasir- Group I**

〔語義一覧：List of senses/meanings〕

1.   人、動物などが、（足を交互にすばやく動かして）速やかに前進する (a person or an animal moves quickly ahead (by quickly moving its legs alternatively))

2.   乗り物が速く動く (a vehicle moves fast)

3.   乗り物が運行する (transportation operates)

4. 目的の場所へ急いで移動する (to move to the destination hurriedly)

5. 目的のために動き回る (to run around for some purpose)

6. 逃げる。自分の立場から逃げてある側につく (to run away, to flee from one's own side and join another side)

7. 好ましくない傾向に傾く (incline towards an undesired trend)

8. 速くさっと見る (take a quick look)

9. 感覚、現象などが一瞬にして現れる（現れて消える）(sudden appearance [and disappearance] of a feeling or phenomenon)

10. 道、川、亀裂などがある方向に延びている、通じている (extension or continuation of a road or a river or a crack etc. in a particular direction)

11. 活動する。実績を上げる (to work, to achieve results)

The details of the sense 1 are described below. Owing to space restrictions, other senses are not discussed here.

〔語義：**Sense/meaning**〕

人、動物などが、（足を交互にすばやく動かして）速やかに前進する
a person or an animal moves quickly ahead (by quickly moving its legs alternatively).

〔表記：**Orthographical form**〕

走（はし）る

〔自他：**Transitivity**〕

自動詞 (Intransitive)

〔イメージ：**Image**〕
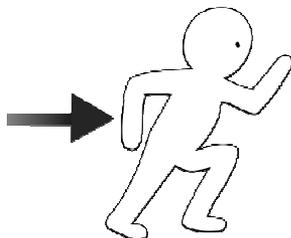
〔構文フレーム：**Construction frame**〕

・ 基本フレーム：<人・動物> が走る (Basic frame: <person/animal> NOM runs)

・ オプショナル要素 (Optional elements/adjuncts)
（起点）から (source) kara,（着点）まで (goal) made
（場所１／位置）を (location 1/position) wo,（距離１）を (distance 1) wo

（場所２）で (location 2) de,（道具）で (instrument) de,（速さ）で (speed) de,
（距離２）(distance 2),（目的）で (purpose) de,（時間１）で (time 1) de,
（様態）(manner),（時間２）(time 2)

〔共起例：Collocations〕

＜人・動物＞が <person/animal> ga	① 人 (person)：私 (I),　ー（さん）Mr./Mrs./Ms. X,　彼 (he),　子供 (child),　選手 (player) ② 動物 (animal)：馬 (horse),　猫 (cat),　ネズミ (mouse)
（起点）から (source/starting point) kara	① 建物 (building)：駅 (station),　家 (house) ② 場所 (place)：東京 (Tokyo),　箱根 (Hakone),　（人／もの）のところ (from the location of a person or an object)
（場所１／位置）を (place 1/position) wo	① 場所 (place)：公園 (park),　屋内 (in-house),　校庭 (school ground),　砂浜 (beach),　〜沿い (along something),　歩道 (walkway),　山道 (mountain trail/pass),　コース (course),　廊下 (corridor),　水の上 (on or above the water surface),　闇の中 (in the dark),　暖かい日差しの中 (in the warm sun) ② 位置 (position)：目の前 (in front of one's eyes),　先頭 (ahead),　トップ (top),　はるか前方 (way ahead in the forward direction)
（距離１）を (distance 1) wo	マラソン (Marathon),　42.195km,　ハーフマラソン (half marathon),　長距離 (long distance),　短距離 (short distance)
（場所２）で (location 2) de	公園 (park),　屋内 (indoor),　校庭 (school ground),　砂浜 (beach)
（道具）で (instrument) de	ジョギングシューズ (jogging shoes),　裸足 (bare foot)
（速さ）で (speed) de	全速力(with full speed),　時速 50km (50 km/hour)
（距離２） (distance 2)	100 メートル (100 meters),　50 メートル (50 meters)
（目的）で (purpose) de	国体 (national tournament),　オリンピック (Olympic),　レース (race)
（時間１）で (time 1) de	1 時間 (one hour),　100 メートルを 11 秒 (100 meters in 11 sec)
（様態） (manner)	ゆっくり (slowly),　速く (fast),　一目散に (as fast as one's legs can/could carry one),　勢いよく(fiercely),　息せき切って (breathlessly),　トロトロ (feebly),　ビュンビュン (zippingly)
（時間２） (time 2)	1 時間 (one hour),　10 分 (10 minutes)

〔非共起例：Wrong collocations〕

（様態）(manner)　　（誤）(inappropriate/incorrect)　　遅く (slowly)

〔例文・作例：**examples/coined examples**〕

- 大学の中を新しい靴でゆっくりと 10km 走った。((I) slowly ran 10 km at the university wearing new shoes.)

- 犬が公園の中を向こうからこちらへ走ってくる。(The dog runs across the park from the other side to here.)

- 駅伝で東京から箱根まで走る。(To run from Tokyo to Hakone in the ekiden race.)

- 駅まで大通りを走っていく。(To run to the station along the boulevard street.)

- 家のまわりをゆっくり 20 分ほど走った。((I) ran slowly around my house for 20 minutes.)

〔例文・コーパス：**examples/from corpus: not translated into the target language**〕

- まるで競争しているみたいな勢いで廊下を走るとは。（ベティ・ニールズ作、和香ちか子訳『幸せへの航海』, 2004）

- ちなみに、お巡りさんは歩道を走っています。(Yahoo!知恵袋, 2005, マナー、冠婚葬祭)

- かなり本格的に走る人たちばかりで、半分ぐらいは外国人の感じもしますが、日本人であれ外国人であれ、こんなに大勢の人が走るのは健康になりたいためでしょうか。(阿久悠『詩小説』, 2000. 9 文学)

〔個別の誤用情報：**Information on errors pertaining to specific use**〕

(1) 語義 1 で到達地点の「に」をとることはできない。到達地点の「に」を用いた場合、語義 4 の解釈になる。語義 1 で到達地点の「に」をとるときは「走っていく」「走りこむ」などと方向を表す動詞を伴った複合動詞にする必要がある。(In the case of sense/meaning 1, the goal location cannot be marked with the particle "*ni*". If the goal location is marked with the particle "*ni*", the meaning changes to sense/meaning number 4. In order to use the particle "*ni*" in the case of sense/meaning 1, it is necessary to use a complex predicate such as "*hashitte iku*" or "*hashirikomu*" which contain a verb implying direction.

（誤: ungrammatical use）駅に全速力で走った。（語義 4 の解釈になる。）
（正: grammatical use）駅まで全速力で走った。
（正: grammatical use）駅に全速力で走っていった。

(2)（様態）の「はやく」は速度を表す「速く」であり、時期を表す「早く」は用いない。The adverb "*hayaku*" is the one that expresses "*speed*" and not the one that expresses "*an early time/period/season*".

（誤: ungrammatical use）あの選手はとても早く走る。
（正: grammatical use）あの選手はとても速く走る。

— 中略 abbreviated —

〔文法：**Grammar**〕

| 語義 sense	走らせる	走ろう	走っている	走った
	使役 causative form	意思 volitional form	進行 progressive form	過去 past form
1	○	○	○	○
2	○	×	○	○
3	○	×	○	△
4	○	△	○	○
5	○	○	○	○
6	△	△	△	○
7	○	×	○	○
8	○	×走らせよう	○走らせている	○走らせた
9	×	×	×	○
10	△	×	○（状態）	×
11	△	△	○（状態）	△

〔複合語： **Compounds**〕

   ▶走り回る　▶走り去る　▶走り通す　▶走り込む　▶走り抜く　▶走り抜ける
   ▶走り過ぎる　▶走り高跳び　▶走り幅跳び　▶突っ走る　▶ひた走る　▶小走り
   ▶ひとっ走り　▶使い走り　▶走り書き　▶走り読み　▶口走る　▶先走る
   ▶才気走る　▶石（いわ）走る　▶血走る　▶ご馳走

〔慣用句・ことわざ: **Idioms/Proverbs**〕

   ▶ペン（筆）が走る　▶虫酸が走る

〔語義ネットワーク: **Semantic network**〕

```
       8 ← 5 → 6 → 7
           ↑
           4
           ↑
10 ← 9 ← ① → 2 → 3
           ↓
          11
```
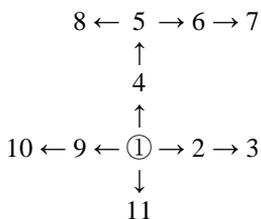
〔関連語（ワードファミリー）: **Related words (word family)**〕

  ・　同義語 Synonyms：▶駆（駈）ける　▶駆け足　▶馳せる　▶ダッシュする

  ・　類義語 Near-synonyms：▶歩く　▶通る　▶動く　▶進む　▶行く　▶飛ぶ

3.2 The methodology of description: the content and the intent

〔**Accent**〕

In the case of accent, H stands for high and L stands for low pitch accent. However, for conveying accent information, the audio medium is more effective than the visual and we provide audio files to convey accent in addition to the visual representation.

〔**Conjugation**〕

The stem of the verb and its conjugation pattern is provided. As for the conjugation pattern, the classification widely used in Japanese language education (Group I, II and III) is adopted.

〔**List of meanings/senses**〕

The basic meaning is presented first and derived meanings follow as distinct senses. The basic meaning is also known as the central meaning and in a polysemous word it is considered as the most basic sense/meaning. This meaning is more concrete, more frequent and corresponds to what is known as the prototypical sense. The order of senses/meanings in the list of senses/meanings is decided taking into consideration the semantic closeness or remoteness of the sense in question to the central meaning. However basically this relationship is not linear there is some inevitable arbitrariness in determination of the order of meaning/senses. A semantic network diagram (described below) is also presented in order to show relationships among meanings graphically.

〔**Meaning/Sense**〕

The meaning/sense is explained in simple, easily understood terms. Some key words are intentionally used in order to make clear the relationships among the explanations. Such a strategy will also help to foster the understanding of connections in the semantic network. Also, the explanation is devised in such a way that the semantic congruence between the constructional meaning suggested by the construction frame discussed later and the core arguments and adjuncts would be easier to comprehend.

〔**Orthographic representation**〕

The orthographic representation in Kanji (Chinese) characters is provided with kana reading.

〔**Transitivity**〕

The transitivity of the verb in question is given. Depending on the meaning/sense, the transitivity may vary. However, the transitivity given here is that of the basic/central meaning.

〔**Image**〕

Providing a pictorial image of the meaning/sense helps in facilitating understanding of the meaning/sense in question. Image plays an important role especially in the derived/extended meanings/senses. Images are modeled on image schema proposed in the theory of cognitive linguistics. However we adopted more concrete images as compared to theoretical image schema. Further, in the case of image, unlike image

schema, emphasis is given to ease of understanding rather than theoretical precision. For the image, still pictures, animation as well as video clips are used (see section 6 for details).

〔**Construction frame**〕

The construction frame is shown in the form of a two tier structure: obligatory core arguments and optional adjuncts. However, as shown below, in some cases judgment between the two is difficult. For example, the verb *kaku* "to write" is a two-place predicate taking two core arguments, however in a construction like <person 1> write <a letter> to <person 2> it behaves like a 3-place predicate. In such cases, in the construction grammar approach (cf. Goldberg (1995), the construction containing 3 arguments is assumed. One falls in a dilemma on the issue of whether the 3-place construction should be incorporated in the description of a dictionary entry for the verb *kaku* "write". This is because, if one proceeds with adopting the construction-centered explanation, one needs to include extremely eccentric constructions as well, resulting in dramatically swelling the length of the description. Even if one adopts such a description policy, the issue of deciding whether the phrase <person 2> *ni* should be treated as an argument or as an adjunct remains unsolved. Viewed from the meaning/sense of the verb it is an adjunct while viewed from the point of a construction it is an argument. At present this issue is left to the decision of the entry writer and editor, however, by referring to the frequency count, this issue can be resolved to a certain extent.

〔**Collocations**〕

Collocations are shown for both arguments and adjuncts. This is because collocations differ from one sense to another as well as from one case particle to another. Collocations are ordered in the sequence of collocation frequency deduced using the BCCWJ corpus browsing tool called NINJAL-LWP for BCCWJ (NLB for short). As a statistical index expressing the strength of a collocation, a score called "Mutual Information (MI)" score is available, however the MI score tends to cull expressions involving high degree of idiomaticity, so we decided to use raw frequency as a criterion for the purpose of listing collocations. Arranging collocation based on the raw frequency deduced from NLB ensures objectivity and authenticity. However, on the other hand, owing to the limitation on the size of the corpus (65 million words in NLB, 100 million words in BCCWJ) there is no guarantee that all the collocations needed to be listed in the dictionary are culled without any leakage. Therefore, some collocations which do not appear in the NLB, but which are thought to be necessary for learners are added. This measure, to a large extent, depends on the experience of the editor. In future, if the size of the corpus is increased, it is expected that the selection of collocations on the basis of the frequency criterion would become easier. For this purpose, the Tsukuba WEB Corpus (TWC) with a projected ten times the entries of BCCWJ is under preparation.

〔**Wrong collocations**〕

Here collocations which are prone to lead to wrong usage are described.

〔**Examples: coined examples**〕

For each meaning/sense we provided more than 3 coined examples. In order to avoid examples ending only with dictionary form (plain style, non-past), we have made a deliberate attempt to coin examples involving variation of tense, aspect, voice, modality etc. Such a move also helps to enhance naturalness of examples. Quite often we have even used complex sentences as well.

〔**Examples: from corpus**〕

We have provided examples culled from the BCCWJ corpus as well. The purpose of providing examples from corpus is to provide examples that are natural in the context of situation in question. However, on the other hand there is the criticism that such examples are difficult for non-natives to comprehend. The same observation has been made during the process of compilation of this handbook as well. It has been pointed out that real examples from a corpus are hard to comprehend unless one has sufficient knowledge of socio-cultural background. It became clear in our handbook that translation of such examples into another language is a big obstacle. Especially, considering the typically High Context Communication (Hall, 1976) nature of Japanese, it is easy to imagine that the problems of real examples would be much graver than in English. Whether to stick to real-examples only or to allow coined examples for the point of view of second language education is a complex issue with no satisfactory solution. At present, taking merits of both, we have decided to include natural examples as well as tailored examples. However, since the translation of natural examples is an extremely difficult task, we have decided not to translate the corpus examples.

〔**Information on wrong usage: in the case of specific meanings**〕

Mistakes that learners tend to make often are described under this heading. For information on wrong usage by JFL learners, various databases including Teramura database (http://teramuradb.ninjal.ac.jp/) are used. However, since these corpora are developed individually, the size of each of them is rather small and it is difficult to deduce general patterns of mistakes from them. Under such circumstances we have to heavily rely on the teaching experience of the editor. The following are examples from learners' corpora:

Spoken language corpus:

発話対照データベース、生活対照データベース (taiwa taishou detaabeesu, seikatsu taishou deetabeesu)

日本語学習者会話データベース(nihongo gakushuusha kaiwa deetabeesu)

日本語学習者会話ストラテジーデータ (nihongo gakushuusha kaiwa sutoratejiideeta)

KY コーパス (KY koopasu)

タグ付き KY コーパスと検索ツール(tagutsuki KY koopasu to kensaku tsuuru)

BTS による多言語話し言葉コーパス (BTS ni yoru tagengo hanashikotoba koopasu)

インタビュー形式による日本語会話データベース（上村コーパス）(intabyuu keishiki ni yoru nihongo kaiwa deetabeesu (Uemura koopasu))

Written language corpus:

寺村誤用例集データベース (Teramura goyou reishuu deetabeesu)

日本語学習者言語コーパス (nihongo gakushuusha gengo koopasu)

作文対訳 DB (sakubun taiyaku DB)

自 然 言 語 処 理 の 技 術 を 利 用 し た タ グ 付 き 学 習 者 作 文 コ ー パ ス (shizengengoshori no gijutsu wo riyou shita tagutsuki gakushuusha sakubun koopasu)

日 本 ・ 韓 国 ・ 台 湾 の 大 学 生 に よ る 日 本 語 意 見 文 デ ー タ ベ ー ス (nihon/kannkoku.taiwan no daigakusei ni yoru nihongoikenbun deetabeesu)

JLPTUFS 作文コーパス(JLPTUFS sakubun koopasu)

In addition to the above list, there are many corpora which are either not made public or are accessible to only few individuals. For the effective use of intellectual resources, it is desired that an organization like NINJAL take the lead in the development of a platform like CHILDES (Child Language Data Exchange System) which allows accumulation of data in a common platform.

〔**Grammar**〕

Here we have shown the behavior of the verb with respect to grammatical categories like aspect, voice, tense etc. A conclusion is still not reached on whether to include categories like direct passive, indirect passive, imperative form, other sentence-final expressions. Further, whether to make judgments on grammaticality of such categories based on intuitions of individuals or on the basis of corpus frequency is also not yet decided. For making judgments on grammaticality (especially the subtle ones, shown by triangle sign) on the basis of corpus frequency, the size of the BCCWJ corpus seems not to be sufficient.

〔**Compounds**〕

Compound words are too large in number and hence it is impractical to include all of them. If so, again one has to decide on the basis either of intuition or of corpus frequency in order to decide potential candidates that should be listed. We would like to make use of the corpus for this and at present are using frequency as a criterion for listing compound words.

〔**Idioms and proverbs**〕

Idioms and proverbs consist of elements which are tightly bound together and the meaning of the whole cannot be guessed from the combination of the meanings of the parts. In other words, it can be said that semantic transparency is low in the case of idioms and proverbs. However, the transparency is a gradient concept and the decision of collocation or proverb is bound to be arbitrary. One yardstick for this decision can be MI (Mutual Information) score. The higher the degree of idiomaticity the greater the MI score (see section 4.1.2).

〔**Semantic network**〕

The relationships among meanings/senses are visually shown with the help of a radial category network diagram. The basic or central meaning is the one that is known in cognitive linguistics as the prototypical meaning. The relationships among meanings/senses are visually shown with the help of a radial category network diagram. The basic or central meaning is the one that is known in cognitive linguistics as the prototypical meaning. Derivations from it are arranged in a way to be understood intuitively. These semantic derivations themselves are products of linguistic research. Many cognitive linguists are also involved in this project. However, there is no guarantee that the semantic derivations are determined on the basis of a single meaning. Also the sequence of diachronic change and synchronic relationship often do not match. In view of these considerstions, while insights from cognitive linguistics form the basis of description, often changes have been made in favour of intuitive understanding. There are places where accuracy of description from the point of cognitive linguistics conflicts with intuitive understanding. In such cases we have preferred educational considerations such as ease of understanding for teachers and learners.

As for the network, showing just the connection is not enough. The strength of the connection should also be shown. We are thinking of showing the strength or weakness of the connections visually in terms of the thickness of the line or the distance between the senses so as to foster understanding in a visual and intuitive way.

〔**Related words (word family)**〕

At present, we have listed words with almost the same meaning and synonyms as related words. Listing of antonyms is also under consideration. We are thinking of presenting the word family in the form of a radial category network, if possible.

4.   Developing tools for corpora of correct usage and wrong usage

One of the important policies we adopted to create this handbook is to make good use of available corpora. To compile a corpus-based handbook or dictionary, the existence of tools which enable dictionary writers to use corpora adequately and efficiently in the process of dictionary making is indispensable. In this project we chose the Balanced Corpus of Contemporary Written Japanese (BCCWJ) as a corpus of correct use by Japanese natives and the *Gaikokujin gakushuusha no nihongo goyoureishuu* (*Collection of errors of JFL learners*, 1990), compiled by Hideo Teramura and his colleageus, as a corpus of wrong usages of JFL learners. We developed search tools for each of these corpora. In the following two subsections, we will describe the features and functions of both tools.

4.1 NINJAL-LWP for BCCWJ (NLB)

NINJAL-LWP for BCCWJ (NLB, http://nlb.ninjal.ac.jp) is an online search tool for the BCCWJ, jointly developed by the National Institute of Japanese Language and

Linguistics (NINJAL) and Lago Institute of Language (LIL). The basic unit of this system is LagoWordProfiler (LWP), which LIL has developed for dictionary writing and editing. LWP has been successfully utilized in several projects of English-Japanese, Japanese-English dictionary making.

Figure 1: The headword Window of NLB

BCCWJ is the first balanced corpus of the Japanese language, developed by NINJAL, and its final version was made public at the end of 2011. It is a large corpus of more than 100 million words, the size of which is comparable to the British National Corpus. The main component of the corpus consists of random samples from books, newspapers, magazines using rigid statistical methods to establish representativeness. Nine additional sub-corpora are provided for special purposes, including web text, which shows different usage patterns from those of text of the print media (Maekawa, 2012).

4.1.1 Lexical profiling

The most important feature of NLB is its introduction of the lexical profiling methodology. Lexical profiling is now a standard method for making corpus-based dictionaries because it satisfies the requirements for using corpora in dictionary making. A concordancer used to be a standard tool in the earliest corpus lexicography. On the COBUILD Project, which made extensive use of corpora for the first time, the writing staff wrote headword entries by analyzing concordance lines from a concordancer (Sinclair, 1987). Concordance lines enable the dictionary writer to analyze individual words in real context. However, the larger the number of lines, the more difficult it is to grasp the whole variety of linguistic phenomena. To solve this difficulty, lexicographers realized the importance of summarizing linguistic phenomena comprehensively by use of abstraction (lemmatization, POS tagging, and chunking) and statistical measures (the MI score, the T score, etc.). In this process,

lexical profiling as a new approach gradually developed (Church et al., 1991). At the end of the 1990s, a practical lexical profiling tool called *Word Sketch* appeared (Kilgarriff & Rundell, 2002). This software was first used for compiling *Macmillan English Dictionary for Advanced Learners*, and then it developed into the integrated system *Sketch Engine*, which is now used in many dictionary projects.

Lexical profiling has two important requirements. The first is comprehensiveness. Linguistic research, in general, focuses on a particular linguistic behavior and adopts an approach that examines individual instances carefully and thoroughly. On the other hand, what dictionary making requires is to examine each headword's overall behavior. A dictionary writer needs to grasp a headword's behavior as comprehensively as possible. When implementing a search tool, which patterns to extract and how to classify those extracted patterns are vital keys to ensure comprehensiveness.

The other key is time efficiency. This is essential in dictionary making. The number of headwords in a dictionary range from several thousand to one hundred thousand. To make best use of a corpus when writing a large number of headwords, an environment that enables dictionary writers to use a corpus efficiently is indispensable. Key factors to realize this environment include search speed and a user interface.

4.1.2 Lexical profiling in NLB

So how does NLB satisfy the requirements of lexical profiling? As to comprehensiveness, NLB deals with the orthographical variety of the Japanese language.

Figure 2: Orthographical forms for *toriageru*

Japanese is usually written in three types of characters: *hiragana*, *katakana* and *kanji*. This means a word could be written in at least three ways. The noun *hito*, which means *a person*, can be written as ひと in hiragana, or ヒト in katakana, or 人 in kanji, with different connotations. In the case of compound verbs, things are complicated by the fact that some verbs have two or more kanji candidates with slightly different meanings. The compound verb 取り上げる (toriageru), which means *pick up* or *adopt*, can also be written as 採り上げる. Including a variation of kana suffixes, more than ten orthographical forms for トリアゲル are possible. From the point of view of comprehensiveness, it is, in many cases, more appropriate to group two or more orthographical variants into the most typical orthographical form than to give each form a headword status. NLB deals with this issue by incorporating the idea of representative orthographical form. In the previous example of 取り上げる, more than ten orthographical forms are all grouped into the

representative form 取り上げる, which consists of a headword entry. Figure 2 shows the frequency distribution of orthographical forms for 取り上げる in BCCWJ.

In order to maximize time efficiency, NLB has a user interface that allows the user to examine grammatical patterns, collocations, and examples from the corpus in the same window (See Figure 1). On *Sketch Engine*, which we mentioned earlier, a screen transition occurs every time the user looks for examples for each collocation. A user interface with frequent screen transitions is problematic from the point of view of time efficiency. With the recent spread of large screen displays, it is not so difficult as before to introduce a user interface with a minimum of screen transitions. Although user interfaces for corpus search tools have not been given much attention until recently, its importance is expected to increase as the size of corpora increases and more sophisticated search functions are implemented. Search speed is another important factor closely related to time efficiency. NLB shows the results of collocations and examples almost instantly by optimizing the structure of the database.

Another important feature of NLB is its function to sort collocations by raw frequency and other statistic measures such as the MI-score and the logDice score. Figure 3 shows collocations of N を買う(*N wo kau*, to buy *N*). In the upper part of the figure, collocations are ordered by raw frequency, and in the lower part, by MI score. The MI score has a tendency to be unreliably high among low-frequency collocations. To avoid this reliability issue, NLB provides a filter function to remove low-frequency collocations. In the lower part of Figure 3, low-frequency collocations of less than five instances are excluded from the list. You can see idiomatic expressions like 顰蹙を買う(upset someone), 歓心を買う(seek someone's favor), 失笑を買う(make someone laugh at you) are top of the list. Sorting collocations by multiple statistic measures is an extremely useful function.

...を買う 1264件					...を買う 180件			
コロケーション	頻度	MI	N-S		コロケーション	頻度	MI	N-S
ものを買う	177	3.92	-2.29		顰蹙を買う	46	13.01	-0.48
本を買う	119	5.72	-0.30		歓心を買う	18	13.01	-0.40
【一般】を買う	118	2.12	-0.26		不興を買う	25	12.69	-0.04
車を買う	99	6.38	0.16		失笑を買う	9	11.69	-0.22
物を買う	96	6.33	0.57		不評を買う	10	11.69	-0.02
株を買う	87	8.42	-1.11		反感を買う	39	10.97	0.14
切符を買う	79	10.77	-0.48		切符を買う	79	10.77	-0.48
家を買う	74	4.43	-0.23		パンプスを買う	6	10.21	0.02
のを買う	71	1.03	-1.99		馬券を買う	21	9.78	0.14
土地を買う	66	6.84	-0.46		土産を買う	45	9.70	0.21
それを買う	59	2.64	-0.91		宝くじを買う	12	9.69	-0.40
服を買う	56	7.85	-0.63		おみやげを買う	9	9.52	-0.04
品を買う	50	6.70	0.12		一役を買う	5	9.46	0.09
商品を買う	49	6.27	0.32		怨みを買う	34	9.42	-1.27
券を買う	46	8.73	0.50		ウーロン茶を買う	5	9.21	0.11
顰蹙を買う	46	13.01	-0.48		チケットを買う	23	9.03	0.16
Page 1 /13 100					Page 1 /2 100			

Figure 3: Collocations of *N wo kau*

NLB also facilitates creating examples with dictionary-making-oriented functionality. On the example panel (the right-most panel of Figure 1), examples for a collocation are shown in ascending order of their character counts. This helps the dictionary writer to use corpus examples for reference easily and effectively. Each corpus example is color-coded according to the sub-corpus it belongs to, which enables the writer to know where each example comes from quickly. In addition, the writer can examine the context of a corpus example just by clicking its source information label.

As we have seen, NLB provides an ideal environment for Japanese dictionary making, by dealing with the wide variety of orthographical forms in Japanese, and offering a user-friendly interface.

4.2 The Teramura Wrong Usage Database

Gaikokujin gakushuusha no nihongo goyoureishuu (*Collection of errors of JFL learners*) is a report compiled by Teramura Hideo and his team in the late 1990s, after they collected and classified misuse samples from compositions written by overseas students from 24 countries. The total of the misuse samples amounts to 6,300, with misuse labels attached to misuse positions. Other information includes learner's nationality and composition type.

The online version of this report, Teramura Wrong Usage Database provides a search function. The user can search misuse examples by combining conditions (a type of misuse, a learner's nationality, a composition type, etc.) Figure 4 shows the "search

from misuse type" function. Misuse types are shown in a tree structure, effectively informing the user of how many misuse instances there are for each type on any combination of nationalities and composition types.
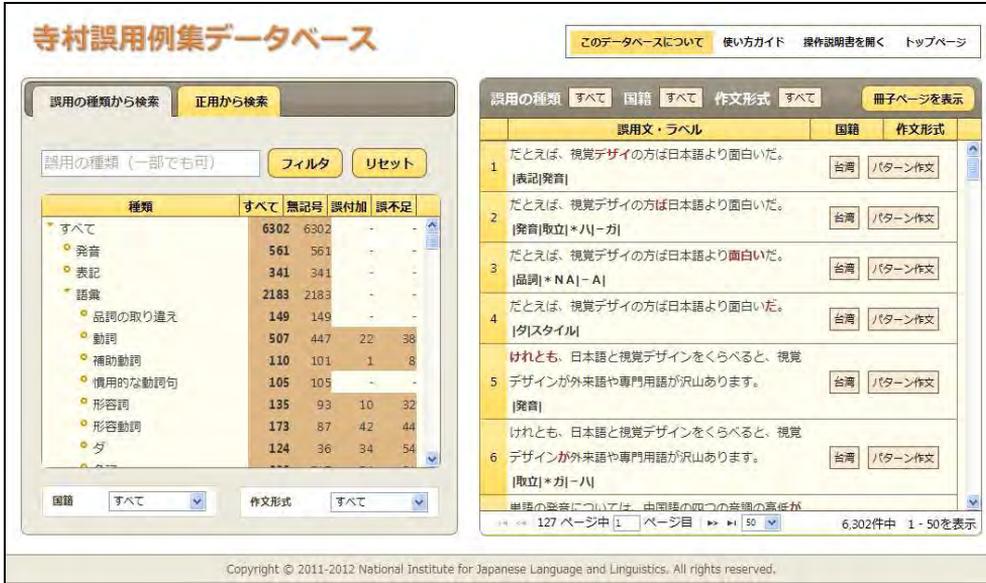
Figure 4: Teramura Wrong Usage Database

Most conventional Japanese dictionaries for native speakers and foreign learners, including ones with a learning or teaching purpose, only show correct usages; very few show wrong usages. This tool enables us to include useful wrong usage information for learners such as wrong collocations in a definition entry.

5. Crossing the barriers of space and time: An online multi-lingual editing tool

Compiling a dictionary requires a lot of time and human resources. It is usually the case that there is an editor-in-chief who directs lexicographers in charge of writing up entries. The editor-in-chief proofreads the entries that the lexicographers have written, and corresponds with them as often as necessary. Proofreading may be done by different proofreaders and the editor-in-chief manages the editorial activity. This process usually takes a long time, and is not ideal if time for the compilation is limited. Another drawback of this traditional system is that lexicographers will usually have no chance to examine entries that the other lexicographers write.

To overcome these problems, we have developed a web-based editing system so that the editors, lexicographers and proofreaders can have access to the entry data for editing, reviewing and proofreading processes.

To develop the current online editor system, our experience in compiling A Dictionary of Basic Verbs in Japanese for Marathi, the outcome of Prashant et al. (2007)'s project is fully exploited. Under a limited budget, we made use of free applications to achieve our goal: a wiki system to store the entry data in XML format. Wiki is a system for collaborative editing online and has a repository system, under which all older versions of wiki pages are stored. By comparing the current version with one of the older versions, editors can tell what have been changed, deleted and/or added in the latest version. In this new system, we take advantage of the repository feature of wiki.

In the current system, the lexicographers write entries in Japanese first. Then the Japanese entries are translated into four foreign languages (Marathi, Korean, Chinese and English) by translators. At this stage some additional information will be added that is related to cultural and linguistic differences between Japanese and the target language.

The following sub-sections give a brief outline of the online editorial system.

5.1 An outline of the online editorial system

5.1.1 Some features of the online editor

The online editor developed for this project has the following features:

- Data are input in a textbox area on the editor and stored in an XML structure.

- The data input in the editor are reflected in a preview function to check how they look in the HTML format instantaneously.

- Employing Yahoo API, it is possible to assign furigana, the phonetic transcription of kanji, in a format that may be convertible into other formats like HTML.

- The lexicographers can read the entries that are written by the others online and post a comment, which will be shared by all editors.

5.1.2 Online editor as a plug-in of Dokuwiki

The editor is not a standalone application but is developed as a plug-in for Dokuwiki. Dokuwiki is a Unicode-based wiki application and does not require a binary database system like SQL because data pages are saved in text files. Each entry is organized in an XML format and stored as a Dokuwiki page. Since the file is a text file, it can be directly used as an XML file for data-processing.

The lexicographers first login to the Dokuwiki homepage as in Figure 5.

Figure 5: The homepage of the editorial system on Dokuwiki

5.1.3 Starting the online editor

After logging in, lexicographers choose the language, and then select one of the entries in the list to edit it. The Wiki page shows the XML data of the entry, but it is not directly edited. They start the plug-in online editor. On starting up, the editor retrieves the XML data from the Wiki page. The view of the entry data is formatted in an Explorer view, with a tree structure displayed on the left pane and each sub-data displayed on the right pane, as in Figure 6.

Figure 6: A full view of the online editor

Figure 7 shows the view when one of the items is selected and its editing area is displayed on the right pane.
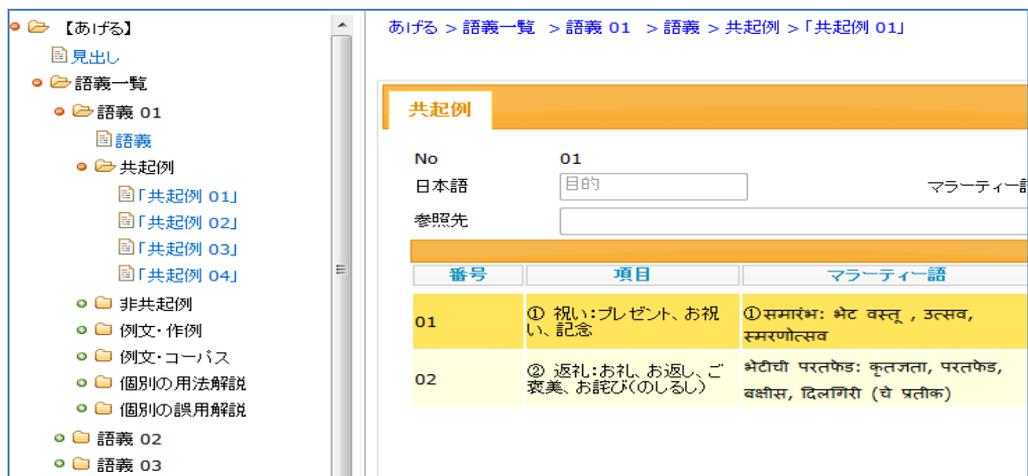
Figure 7: The editing pane for Collocation 01（共起例 01） is open on the right page

5.1.4 The preview function

The editor has a preview function. There are two types of preview: the entire view of the entry and the partial view of an item of the entry. The preview is generated via XSLT as an HTML page. An image of the full-scaled preview in Marathi is shown in Figure 8, and an image of the partial view is shown in Figure 9. Since it is a bilingual version, both the Japanese data and the respective Marathi data are shown. In the bilingual version, as shown in Figure 8, an additional piece of information from a contrastive point of view（対照情報） is also provided when necessary. This information will not be included in the Japanese version.
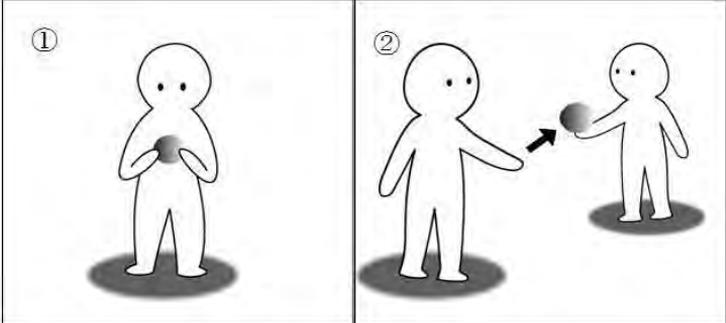
Figure 8: The full-scaled preview of the Marathi translation of *ageru*

The layout design of the preview in Figures 8 and 9 is not intended to be final, but to be temporary just for convenience. The final layout design will be developed differently and be applied to generate the final product from the same XML data.
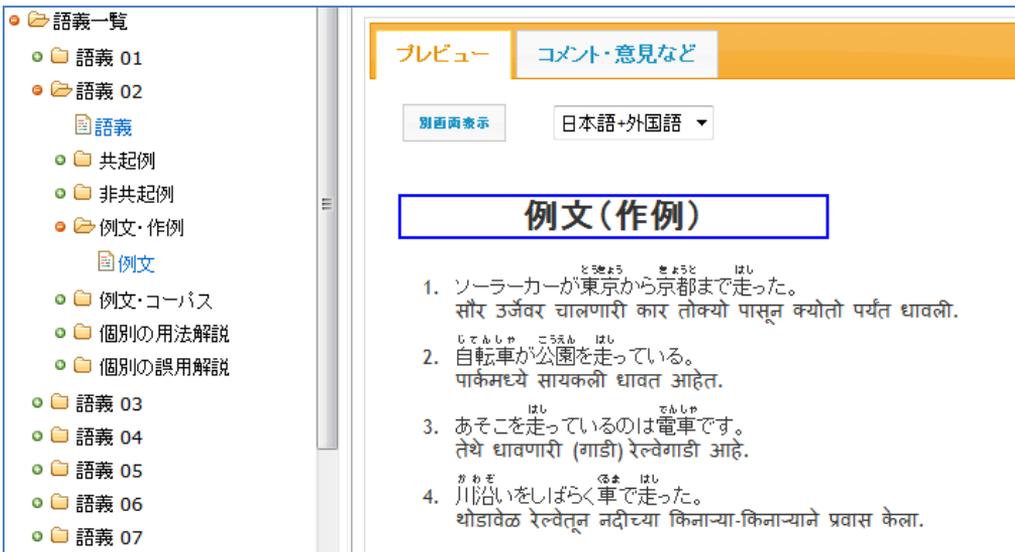
Figure 9: Built-in partial preview of portion of examples

5.1.5 Comparison of different versions

Dokuwiki's revision control makes it possible to compare the latest version with any older version. When two versions are compared, differences will be displayed. This is one of the major merits in using Dokuwiki for entry data management.

5.1.6 Posting comments and improving the quality of description

The editor has a function of posting a comment on the data, while editing or reviewing. Comments are sent to all the editorial members to share the information by email. The comments can also be viewed on the editor and follow-up comments posted. Through this process, the editorial members can exchange ideas and opinions about entry data so that the lexicographer in charge can improve the quality of the descriptions and examples.

6.   Audio-visual contents

Taking into consideration the cognition and memorization process in learning new words and meanings, we have incorporated audio-visual contents in the handbook. We believe audio-visual contents facilitate the comprehension and memorization of various meanings of a verb. A brief discussion of the audio-visual contents is provided below.

6.1 Audio-visual contents

In the present handbook, before presenting specific meanings of a polysemous verb, we first provide an abstract image schema which represents the core, shared meaning of the verb in question. Following this, a radial semantic network of the various meanings of a polysemous verb is provided. These two visual contents set the stage for zooming into a specific meaning. On moving to a specific meaning, we provide an animated illustration of the representative example of that meaning. The animated illustrations are a set of still hand-drawn pictures which are connected in such a way that they depict the semantic scenario as it unfolds in time. Audio contents are also added to the animated illustrations. In addition to the abstract image-schema animated illustration depicting a specific meaning, we also provide video-clips as well. A brief description of these three audio-visual contents is given below.

6.1.1 Image schema

The verbs included in the present handbook are all fairly polysemous. For example, the entry of *agaru* "rise/go up" in our handbook has as many as 19 meanings. In the cognitive linguistics perspective all these meanings are considered to share a core or prototypical meaning which is illustrated with the help of an abstract line diagram which is widely referred to as an image schema. For example, in the case of

the verb *agaru* "rise/move up" "motion of an entity in a upward direction" is taken to be the core meaning and it is illustrated with the help of an image schema shown in figure 10.
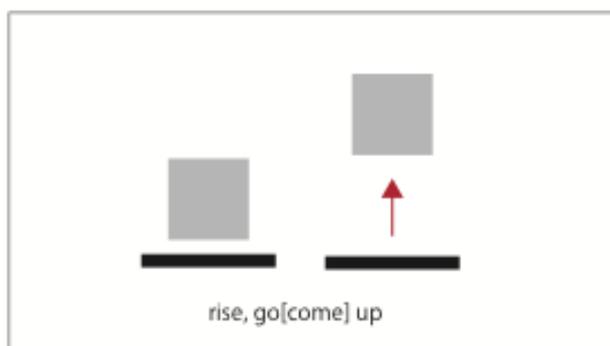

Figure 10: Image schema for the verb *agaru* "rise/move up"

All the meanings are derived from this prototypical meaning through semantic extensions of various types. Image schema would be useful for learners to understand the core or prototypical meaning of a polysemous verb and also to appreciate the connection with specific meanings.

6.1.2 Animated illustrations

As mentioned earlier, we have included animated illustrations in order to facilitate comprehension and retention of specific meanings. From previous research (Dwyer, 1978; Lin & Dwyer, 2004; Dwyer, 2007; Chou & Hsiao, 2010), it has become increasingly clear that the static visual instruction serves as a powerful learning strategy and improves information acquisition and retrieval capabilities. Using these insights, rather than using multimedia contents, we use multiple hand-drawn animations and we show them in a sequence along with audio contents synchronized with them as shown in Figure 11.


Figure 11: Animated illustrations for *ie ni agaru* "to visit someone"

6.1.3 Video clips

Contextbased information such as deixis (e.g. the use of auxiliary verb indicating the location of the speaker in expressions such as *agatte iku/kuru* (come/go up)), or the resultant state conveyed by *-te iru-* form as in the expressions such as *hata-ga agatte iru* (the flag is raised), or the perfectly fried, crisp *tenpura* as depicted by the adverb *karatto* and the like can be effectively conveyed using a video clip. Wherever necessary and feasible, we have tried to provide video-clips to foster understanding of subtle meanings. Figure 12 below illustrates the expression *kaidan o agatte iku* (go up climbing the stairs) wherein the scenario is shot from the backside of the person climbing the staircase to induce the viewers "viewpoint" in the interpretation of the scene.
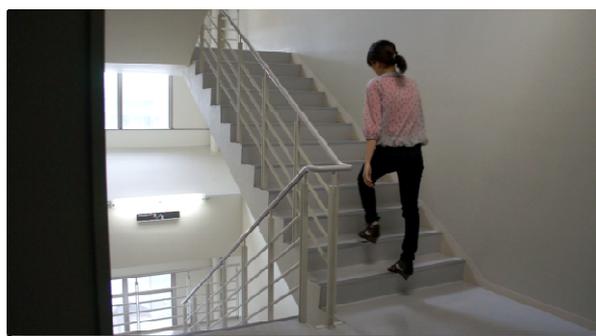
Figure 12: Video clip depicting *kaidan o agatte iku* (go up climbing the stairs)

7.   Future prospects

From the foregoing discussion it should be clear that the handbook in preparation differs in many respects from bilingual dictionaries available now. The content of the entries is based on information gleaned from corpora and is augmented with insights from various sub-fields of linguistics, especially cognitive linguistics and contrastive linguistics. Further, the handbook includes audio-visual contents in order to improve information acquisition and retrieval capabilities.

The handbook will be made available for free access on internet around April 2013. After getting feedback from JFL learners and teachers of Japanese in various parts of the globe, we plan to make improvements both in content as well as presentation. We also plan to increase the number of headwords and the target languages beyond English, Chinese, Korean and Marathi, if collaborators are willing to volunteer their services. Finally, we strongly believe that the output of the handbook project will make a substantial contribution not only to Japanese language research and Japanese language pedagogy but also to corpus linguistics, contrastive linguistics and linguistics in general.

References

Church, K., Gale, W., Hanks, P. & Hindle, D. (1991). Using Statistics in Lexical Analysis. In: Sernik, E. (Ed.) *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, 115-164. New Jersey: Psychology Press.

Chou, P., & Hsiao, H. (2010). The Effect of Static Visual Instruction on Students' Online Learning: A Pilot Study. *Interdisciplinary Journal of Information, Knowledge, and Management.* 5. [Available online: http://www.ijikm.org/Volume5/IJIKMv5p073-081Chou456.pdf]

Dwyer, F. M. (1978). *Strategies for improving visual learning*. State College, PA: Learning Services.

Dwyer, F. M. (2007). The program of systematic evaluation (PSE): Evaluating the effects of multimedia instruction 1965-2007. Educational Technology, XLVII(5), 41-45.

Goldberg, A. F. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

Hall, E. T. (1976). *Beyond Culture*. Anchor books.

Kilgarriff, A., & Rundell, M. (2002). Lexical Profiling Software and its Lexicographic Applications: A Case Study. In: Braasch, A. & Povlsen, C. (Eds.) *Proceedings of the Tenth EURALEX Congress*, 807-819. Copenhagen.

Lin, C., & Dwyer, F. M. (2004). Effect of varied animated enhancement strategies in facilitating achievement of different educational objectives. *International Journal of Instructional Media*, 31(2), 185-198.

Maekawa, K. (2012). Gendai kakikotoba kinkou koopasu (BCCWJ) no kouchiku to KOTONOHA keikaku no ayumi (The construction of "the Balanced Corpus of Contemporary Written Japanese (BCCWJ)" and the progress of the KOTONOHA plan). *Nihongengogakkai dai 144 kai taikai yokoushuu* (Proceedings of the 144th meeting of the Linguistic Society of Japan), 352-357.

Pardeshi, P., & Akasegawa, S. (2011). BCCWJ wo katsuyou shita kihondoushi handobukku sakusei: koopasu buraujingu shisutemu NINJAL-LWP no tokuchou to kinou (Compilation of basic verbs handbook using the BCCWJ corpus: Salient features and functions of the corpus browsing system NINJAL-LWP). *Gendai kakikotoba kinkou koopasu (BCCWJ) kansei kinen kouenkai yokoushuu* (*The proceedings of the symposium commemorating the completion of "the Balanced Corpus of Contemporary Written Japanese (BCCWJ)"*), 205-216. Tokyo: National Institute for Japanese Language and Linguistics.

Pardeshi, P., & Kiryu, K. (2007). "Nihongo-Maraathiigo kihondoushiyouhoujiten" sakusei purojekuto: indo ni okeru nihongo kyouiku no kisozukuri ni mukete ("Japanese-Marathi Basic Verb Dictionary" Compilation Project: A step toward the construction of the foundation of Japanese language education in India). *Koube daigaku ryuugakusei sentaa kiyou* (*Bulletin of Kobe University International Student Center*) 13: 87-102.

Sinclair, J. (1987). *Collins Cobuild Dictionary English Language Dictionary*. London: Harper Collins Publishers.

Sinclair, J. (Ed.) (1987). *Looking Up, An account of the COBUILD Project in lexical computing*. London: Collins ELT.

Teramura, H. (1990). *Gaikokujin gakushuusha no nihongo goyoureishuu* (*Collection of errors of JFL learners*). Report of grant-in-aid study [available online at: http://teramuradb.ninjal.ac.jp/teramura.goyoureishu.pdf].