

# A SURVEY OF REGISTER LABELLING IN JAPANESE DICTIONARIES: TOWARDS A BETTER LABELLING FOR LEARNERS OF JAPANESE

**Kanako MAEBO\***

Hitostubashi University

xiangcai2@gmail.com

## Abstract

Writing by learners of Japanese as a foreign language often contains words that do not fit the style of their context. One possible reason for this is the lack of information on word connotation and usage labels in existing dictionaries. The present study examines the current state of connotational information and register labels in Japanese learner's dictionaries, Japanese language dictionaries and dictionaries of synonyms, and proposes a possible technique for analysing the words' descriptions. The study reveals that Japanese learner's dictionaries and dictionaries for Japanese native speakers use different register labels and assign them from a different perspective. In the case of synonyms, presenting them in their context of use appears to be more useful than only listing them. Finally, the *Balanced Corpus of Contemporary Written Japanese* (BCCWJ - *Gendai Nihongo Kakikotoba Kinkoo Koopasu*) is used to show how corpora can be a very important linguistic resource for the analysis of lexical register.

## Keywords

register; style; Japanese learner's dictionary; Japanese language dictionary; corpus

## Izveček

Izdelki piscev, ki se učijo japonščino kot tuj jezik, pogosto vsebujejo besede, katerih slog se ne ujema s slogom sobesedila. Eden od možnih razlogov za to je pomanjkanje oznak o rabi in informacij o konotacijah besed v obstoječih slovarjih. Članek analizira informacije o konotacijah in oznake o rabi v učnih slovarjih japonščine kot tujega jezika, enojezičnih slovarjih japonščine in slovarjih japonskih sinonimov. Ugotavlja, da slovarji za učence japonščine kot tujega jezika in slovarji za govorce japonščine kot maternega jezika uporabljajo različne oznake in označujejo besede z drugačnih zornih kotov. Opisi sinonimov, ki vključujejo tudi sobesedila, v katerih se pojavljajo, se izkažejo za bolj koristne od golih seznamov. Analiza s pomočjo uravnoteženega korpusa sodobne pisane japonščine (BCCWJ) pokaže, kako so lahko korpusi zelo pomemben jezikovni vir pri analizi leksikalne zvrstnosti.

## Ključne besede

zvrstnost; slog; učni slovar japonščine; slovar japonščine; korpus

---

\* Translated by Jasmina Krofič and Kristina Hmeljak Sangawa

## 1. Introduction

The writing of learners of Japanese as a foreign language (hereafter referred to as learners) often exhibits problems at various levels, including grammar, vocabulary or expression, composition and argumentation. The choice between colloquial or typically written language expressions which fit the style of writing is often a problem even for advanced level learners.

The cause of this lies in the specific characteristics of the Japanese language. Japanese is considered to be a language which strictly distinguishes between written and spoken language. According to Nakamichi (1989) “in Japanese there is an extremely high demand to maintain a fixed style within a text. The stylistic value of a word sometimes exhibits as much regulatory force as its semantic value. Furthermore, it may even happen that stylistic suitability is given priority over grammatical correctness or appropriate meaning.”<sup>1</sup>

The major problem learners face here is that despite these rigorous restrictions on the style of texts and words, information on the stylistic characteristics of words readily available for learners is limited. While producing texts, learners make use of dictionaries to investigate word meaning and usage. However, judging from learners’ writings, it is often clear that they did not get enough information about the stylistic peculiarities of the words they use. Their knowledge of a word’s stylistic characteristics is usually acquired through lists of differences between written and spoken language which can be found in textbooks on writing for foreign language learners, or from their teachers’ instructions during composition classes.

This is hardly a desirable state of affairs for learners and teachers. The compilation of a learner’s dictionary enabling the learners to acquire sufficient information regarding a word’s stylistic characteristic is therefore highly desirable.

## 2. The stylistic characteristics and register (*isō* 位相) of words

A word’s stylistic characteristics are “those characteristics held by each particular word which form the style of an entire text”<sup>2</sup> (Miyajima, 1972). Miyajima (1977) divides words into three classes, with the pivot class of stylistically neutral “everyday words” (*nichijōgo* 日常語), alongside which exist “written language words” (*bunshōgo* 文章語) and “slang words” (*zokugo* 俗語). Each of these classes is further divided into subclasses.

---

1 日本語においては、文章を通して一定の文体が維持されることは非常に強い要求であって、語の文体的特徴は、時として語義的特徴と同等の規制力を発揮することがある。さらには、文法的・意味的な正しさよりも文体的なふさわしさが優先されることさえある。

2 文章全体としての文体をなりたさせるような、個々の単語の持っている特徴のこと

The stylistic characteristics of words have also been described as one aspect of *linguistic phase* (*isō* 位相 in Japanese) or register. According to Yonekawa (2002) a linguistic phase (*isō*) refers to the speaker or writer's social attributes (such as age, occupation, gender etc.), the mode of expression and communication used, the setting, the speaker or writer's psychological and language awareness etc.<sup>3</sup>. Considering this definition, we may say that a word's stylistic characteristics are those aspects of a word which are focused on and which become evident in different settings and modes of expression.

When considering the use of words which fit the style of their context, it is necessary to take into consideration who the speaker or writer is and in what environment the words are being used. In the following sections, the word *register* shall be used as an equivalent of the Japanese term *isō* (“*phase*”) to refer to this aspect of word use.

### 3. Aim and research method

The present paper aims at describing the present state of register information in existing dictionaries, and discuss methods that can be used to acquire such information. It reports on a survey of register information in learners' dictionaries, and on a second survey of register information in dictionaries targeted at Japanese native speakers.

In order to describe register information in existing dictionaries, the following dictionary elements are considered and analysed: editorial policy, explanatory notes and directions for use in the dictionary front- or back-matter, lists of symbols and abbreviations used in the dictionary, and mentions of register within dictionary entries. Four pairs of adverbs were selected from lists of expressions highlighting differences between “spoken expressions” and “written expressions” found in textbooks on writing reports and research papers targeted at foreign learners of Japanese. The eight adverbs are: *tabun* [たぶん “perhaps”] and *osoraku* [おそらく “probably”], *zenzen* [全然 “(not) at all”] and *mattaku* [まったく “entirely”], *zenbu* [全部 “all, wholly”] and *subete* [すべて “completely”], *ichiban* [一番 “most”] and *mottomo* [最も “most”]. These adverbs are presented in the textbooks as synonyms, but to be used in different types of texts. They were therefore chosen as the object of this analysis, because it is reasonable to expect that they are words necessitating some information on register in their lexicographical description.

---

3 表現主体の社会的属性(年齢、職業、性別等)、表現主体が使用する表現様式、伝達様式、場面、表現主体の心理・言語意識など

## 4. Results of the surveys

### 4.1 Survey of learners' dictionaries

For English, there are learners' dictionaries based on large scale corpora, which present the senses of each headword in order of frequency, and include common collocations, style labels and plenty of other information that is deemed useful to learners of English as a second language. In contrast, there are only very few learners' dictionaries of Japanese produced in Japan which can be of use to learners above the intermediate or advanced level. Dictionaries with at least 10,000 headwords include *The Kenkyusha's English-Japanese Japanese-English Learners' Pocket Japanese-English Learners' Dictionary – Kenkyūsha einichi nichiei poketto jiten* [研究社 英日・日英ポケット辞典] published by Kenkyusha in 1996 and reprinted in 2009, *Kodansha's Furigana Japanese Dictionary - Furigana eiwa-waei jiten* [ふりがな和英・英和辞典] published by Kodansha in 1999 and reprinted in 2008, the *Informative Japanese Dictionary - Nihongo o manabu hito no jiten* [日本語を学ぶ人の辞典 “Dictionary for people learning Japanese”] published by Shinchosha in 1995, and *Tuttle Concise Japanese Dictionary - Tuttle konsaisu eiwa –waei jiten* [タトル・コンサイス英和・和英辞典] published by Tuttle in 2008. Considering that the receptive vocabulary of adult native speakers of Japanese is considered to be in the range of 40,000 to 50,000 words (Nihongo-kyōiku Gakkai, 1987, p. 295), the size of these dictionaries is clearly unsatisfactory for advanced learners.

Let us now consider to what extent information on word register can be obtained from these dictionaries. Of the dictionaries mentioned above, the following three have labels indicating register, and were further analysed in detail.

- 1) Kenkyusha's English-Japanese Japanese-English Learners' Pocket Dictionary (hereafter referred to as Kenkyusha's)
- 2) Kodansha's *Furigana Japanese Dictionary* (hereafter referred to as *Furigana*)
- 3) *Informative Japanese Dictionary* (hereafter referred to as *Informative*)

*Kenkyusha's* and *Furigana* are Japanese-English bilingual dictionaries, and they include both a Japanese-English and an English-Japanese dictionary bundled together, making it possible to look up words starting from either language. *Informative* is a bilingualised dictionary with Japanese headwords, definitions and examples in Japanese, and English and Chinese translations of most headwords.

The register labels used in these dictionaries are summarised in Tables 1 and 2.

**Table 1:** Register labels used in three learners' dictionaries  
(dots ○ indicate that the label is used, minus-signs indicate that the label is not used)

	<i>Kenkyusha's</i>	<i>Furigana</i>	<i>Informative</i>
Formal	○	○	—
Informal	○	—	—
semi-formal	—	○	—
colloquial	○	○	—
Honorific	○	○	—
Humble	○	○	—
Polite	○	—	—
Crude	—	○	—
Rude	○	—	—
Brusque	○	—	—
Literary	○	—	—
hanashikotoba [話しことば - "spoken language"]	—	—	○
kakikotoba [書きことば - "written language"]	—	—	○

**Table 2:** Definitions of register labels in the three learners' dictionaries

Label	Explanation
Formal	[ <i>Kenkyusha's</i> ] a word used in formal official situations. [ <i>Furigana</i> ] Words marked as formal are characteristics of formal situations and are not likely to be used in casual conversation. This category includes predicate words that are more polite than semi-formal.
Informal	[ <i>Kenkyusha's</i> ] a word used in relaxed and friendly situations
semi-formal	[ <i>Furigana</i> ] This label refers to predicate words in what the Japanese refer to as ですます体 [ <i>desu-masu-tai</i> , "desu-masu style"]. The semi-formal style, in contrast to the informal style, expresses politeness toward the person(s) the speaker is addressing.
Colloquial	[ <i>Kenkyusha's</i> ] an informal word used in conversation. [ <i>Furigana</i> ] Words marked as colloquial are characteristic of casual conversation and not likely to be used in formal situations.
Honorific	[ <i>Kenkyusha's</i> ] a word indicating respect for others [ <i>Furigana</i> ] This label is used for two types of words. One type is predicate words which express respect for someone by honoring the subject of a sentence. The other type is nouns that express respect for their referents.

Label	Explanation
Humble	[ <i>Kenkyusha's</i> ] a word indicating humility. [ <i>Furigana</i> ] This label is used for two types of words. One type is predicate words which express respect for someone else by humbling the subject of a sentence. The other type is nouns that express respect for someone else by humbling their referents.
Polite	[ <i>Kenkyusha's</i> ] a polite word
Rude	[ <i>Kenkyusha's</i> ] a potentially impolite or offensive word.
Crude	[ <i>Furigana</i> ] Words marked as crude are felt to be inappropriate in polite conversation.
Brusque	[ <i>Kenkyusha's</i> ] a potentially rough or abrupt word.
Literary	[ <i>Kenkyusha's</i> ] a word used in the written language.
<i>hanashikotoba</i> [話しことば “spoken language”]	[ <i>Informative</i> ] おもに話しことばとして使う。[Omo ni hanashi kotoba to shite tsukau - “Mainly used as spoken language”]
<i>Kakikotoba</i> [書きことば “written language”]	[ <i>Informative</i> ] 論文、レポートなどに使う。日常会話などは使わないもの。[Ronbun, repōto nado ni tsukau. Nichijō kaiwa nado wa tsukawanai mono - “Used in essays, reports etc. Not used in daily conversation”]

As can be seen in Table 2, *Kenkyusha's* and *Furigana* both use the labels *formal*, *informal/semi-formal*, *honorific*, *humble*, *colloquial* and *rude/crude*; *Kenkyusha's* also uses three additional labels, *polite*, *brusque* and *literary*, while *Informative* only uses two labels, *spoken* and *written*. These are labels for words which require some consideration regarding the setting in which they are used and the relationship between speaker, listener and referent. Considering the fact that honorific language and expressions indicating the speaker's respect for the listener or referent are very developed and frequent in Japanese, labels such as “honorific” and “humble” are indispensable in a dictionary. However, in these labelling schemes surprisingly little attention is given to words used in writing. Despite this information being as much important, only two of the three dictionaries surveyed above use any label for words used in writing, *literary* in *Kenkyusha's*, and *kakikotoba* (“written language”) in *Informative*. The labels *formal* and *informal/semi-formal*, judging from their explanation, seem to be used for words used in spoken rather than written interactions.

Label explanations are most exhaustive in *Furigana*. Such explanations include the information needed by learners about the situations in which the words are actually used. However, as no examples are given, it is difficult to understand to what type of words these labels are applied. Examples would be helpful not only from the standpoint of learners, but also for teachers, to grasp the range of vocabulary indicated by each label.

In the following paragraphs, dictionary entries are analysed to observe whether labels are used to differentiate between synonyms used in different settings. Dictionary entries for the four pairs of adverbs listed in section 3 were found not to contain the information necessary to know which word should be used in what context, despite the

fact they clearly belong to different registers. Semantic descriptions were also unhelpful from this point of view, as can be seen in the following example in Table 3.

**Table 3:** Dictionary entries for *ichiban* and *mottomo* in Furigana  
(transliteration in square brackets added by present author)

<p>一番 [<i>ichiban</i>] the most (indicating a superlative) [→最も [<i>mottomo</i>]]          日本で一番高い山は富士山です。          [<i>Nihon de ichiban takai yama wa Fujisan desu.</i>]          The highest mountain in Japan is Mt. Fuji.</p>
<p>最も [<i>mottomo</i>] the most (indicating a superlative) [→一番 [<i>ichiban</i>]]          水泳は最もおもしろいスポーツだと思う。          [<i>Suiei wa mottomo omoshiroi supōtsu da to omou.</i>]          I think swimming is the most interesting sport.          エベレストは世界で最も高い山です。          [<i>Eberesuto wa sekai de mottomo takai yama desu.</i>]          Everest is the highest mountain in the world.</p>

Table 3 shows parts of the entries for the words *ichiban* and *mottomo* in *Furigana*. Each entry is accompanied by a cross reference to their respective synonym. However, as the meaning definition is exactly the same and examples are exceedingly similar, these entries give no clue to the difference in register between *mottomo* and *ichiban*. Cross references to synonyms should be accompanied by some information on the differences between them, not only with regard to meaning, but also register. Usage examples should also be chosen or edited to indicate that words are used in different circumstances.

This survey of Japanese learner's dictionaries thus revealed that register labels are present, but that the information learners actually need, regarding the concrete situations of use, is unsatisfactory.

## 4.2 Survey of dictionaries for native speakers of Japanese

As there are currently hardly any dictionaries for learners of the Japanese language above the intermediate level, many foreign language learners in Japan use dictionaries designed for Japanese native speakers. Register labelling was therefore also surveyed in Japanese monolingual dictionaries and thesauruses.

### 4.2.1 Japanese dictionaries

Register labels found in 20 general dictionaries targeted at native speakers of Japanese, including concise monolingual dictionaries and thesauruses, were classified into the following 11 categories of labels: “archaism” [古語 *kogo*], “elegant, refined word” [雅語 *gago*], “written language” [文章語 *bunshōgo*], “spoken language” [口語 *kōgo*], “colloquial/slang” [俗語 *zokugo*], “jargon” [隠語 *ingo*], “vulgarism” [卑語 *higo*], “dialect” [方言 *hougen*], “male language” [男性語 *danseigo*], “female language” [女性語 *joseigo*] and “baby talk” [幼児語 *yōjigo*], as reported in Maebo (2009). The results of this survey seem to indicate at first sight that many more labels are used in dictionaries for native speakers than in learners’ dictionaries, but not all of these labels are used in all dictionaries; each dictionary surveyed contained on average 4 to 7 labels.<sup>4</sup> However, factors such as the lack of clear definitions of the registers mentioned, and the use of different labels for the same words in different dictionaries, make it difficult for learners to find the information they need regarding register in dictionaries targeted at native speakers. These dictionaries probably do not include such information because it may be considered to be something known by any adult native speaker of Japanese.

Let us then consider dictionaries targeted at younger native speakers of Japanese. Three dictionaries targeted at junior-high school students were surveyed:

- 1) *Benesse shinshū kokugo jiten* [ベネッセ新修国語辞典] (“*Benesse’s Japanese Dictionary New Edition*”, 2nd Ed., 2012),
- 2) *Gakken Gendai hyōjun kokugo jiten dai 2 han* [学研現代標準国語辞典第2版] (“*Gakken Contemporary Standard Japanese Dictionary 2nd ed.*”, 2011) and
- 3) *Sanseido’s Reikai shin kokugo jiten dai 8 han* [例解新国語辞典第8版] (“*New Japanese dictionary with examples, 8th ed.*”, 2012).







Register information contained in these dictionaries is discussed below.

---

4 Some of the terms in the above list of labels were not used as labels, but information on register was included in entry definitions. Such information can be equally of help to learners.



1) *Benesse shinshū kokugo jiten* [ベネッセ新修国語辞典]**Table 4:** Register labels and explanations in *Benesse shinshū kokugo jiten*

<p>位相の注意(使う場面と相手に注意することば)</p> <p>使う場面と相手を考え、注意して使ったほうがよい語に次のマークを付けました。</p> <p> 友人同士でおしゃべりをするときになどには使えるが、目上の人や文章に使わない、くだけたことば 例) そいつ</p> <p> 友人同士でも使わないほうがよい、あまりよくないことば 例) ぐる</p> <p> 現在はあまり使われない、古い感じのすることば。文章につかわれることがある。例) 今宵</p>
<p>Translation:</p> <p>Register notes (for words that should only be used paying attention to the situation and to the person addressed)</p> <p>We added the following marks to words that should be used with care, considering the situation and the person to whom one is talking.</p> <p> Informal words that can be used when talking between friends and the like, but not towards higher ranking persons nor in writing, e.g. <i>soitsu</i>.</p> <p> Not very good words that should better not to be used even among friends, e.g. <i>guru</i>.</p> <p> Old-fashioned words that are not used much at present. They are sometimes used in writing.</p>

2) *Gakken Gendai hyōjun kokugo jiten dai 2 han* [学研現代標準国語辞典第2版]**Table 5:** Register labels and explanations in *Gakken gendai hyōjun kokugo jiten dai 2 han*

<p>用法...ことばの使い方の注意点。たとえば次のような注意点があります。</p> <ul style="list-style-type: none"> <li>・「文学的」...主に文学作品、詩歌で使われることば</li> <li>・「文章語的」...主に書くときに使われることば</li> <li>・「俗な言い方」...仲間うちでよく使われる俗語や隠語など</li> <li>・「くだけた言い方」...主に話しことばとして使う</li> </ul>
---

Translation:

Usage... Notes about the usage of words. For example, the following notes are used.

- “literary” ... words used mainly in literary works and poems.
- “written” ... words mainly used when writing.
- “slang” ... slang and jargon frequently used among friends
- “informal language” ... mainly used in spoken language

Other labels are not listed in the legend above, but can be found in the entries themselves, such as “rather old expression” [やや古い言い方 *yaya furui iikata*], “old expression” [古い言い方 *furui iikata*], “formal expression” [改まった言い方 *aratamatta iikata*], “rough, rude expression” [ぞんざいな言い方 *zonzai na iikata*].

### 3) Sanseido’s *Reikai shin kokugo jiten dai 8 han* [例解新国語辞典第8版]

In Sanseido’s *Reikai shinkokugo jiten* (Reikai’s new Japanese dictionary, 8th ed.), there are no special notes on labels. However, headword explanations include phrases such as “formal expression for ...” [一のあらたまった言いかた ... *no aratamatta iikata*], “formal word for ...” [一の形式ばったことば ... *no keishikibatta kotoba*], “old-fashioned word for ...” [一の古めかしいことば ... *no furumekashii kotoba*]. There are also notes labelled “phrase” which includes supplementary explanations about the circumstances of usage.

As could be seen in the above survey of label explanations in the three dictionaries for junior-high school students, register labels found in these dictionaries were similar to and not more numerous than those found in dictionaries for adult speakers of Japanese. In other words, labels in dictionaries targeted at younger users also use a typology corresponding to Miyajima’s (1977) strata of lexical style characteristics and are not assigned from the point of view of the concrete situation and medium of use. However, in *Shinshuu kokugo jiten*, register information does consider the circumstances of use and the person addressed, which is a point of view to be considered when labelling words according to register.

Let us now see what type of information is offered regarding the four pairs of adverbs listed in section 3. The entries for these headwords in the three dictionaries for Japanese junior-high school students contained hardly any information regarding the different use of these synonyms. There were only explanations regarding the fact that the adverb *zenzen* [全然 “(not) at all”], which is only used with predicates in the negative form in standard Japanese, is also used with affirmative predicates with the meaning “very” in casual colloquial language.

Since junior-high school students only rarely write short research papers or reports, information to distinguish between words typically used in academic writing and other everyday words could be considered unnecessary. However, dictionaries

targeted at adult users, which are also generally used by high-school students, do not carry such information either, supposedly because it is superfluous for adult native speakers of Japanese who already know how to distinguish between written and colloquial language. In other words, information about which words to use when writing cannot be found in Japanese dictionaries aimed at junior-high school students nor in general dictionaries for adult speakers of Japanese.

#### 4.2.2 Thesauruses

Thesauruses are helpful for understanding the proper use of synonyms. Four Japanese thesauruses were surveyed for register information, and the following labels were found.

- ① *Kadokawa ruigo shin jiten* [角川類語新辞典] (“Kadokawa new dictionary of synonyms”), Kadokawa [角川] 1981/2002.

Register labels included: “literary” [文語 *bungo*], “elegant, refined” [雅語 *gago*], “written language” [文章語 *bunshōgo*], “everyday word” [日常語 *nihijōgo*], “spoken language” [口語 *kōgo*], “slang” [俗語 *zokugo*], “jargon” [隠語 *ingo*], “dialect” [方言 *hōgen*], “old-fashioned expression” [古風な表現 *kofūna hyōgen*], “male language” [男性語 *danseigo*], “female language” [女性語 *joseigo*], and “baby talk” [幼児語 *yōjigo*]. No notes within definitions or descriptions.

- ② *Tsukaikata no wakaru ruigo reikai jiten shinsouhan* [使い方のわかる類語例解辞典新装版] (“*Dictionaries of synonyms with examples to understand their use, new edition*”), Shōgakukan [小学館] 2003.

No register labels. Each dictionary entry describing a group of synonyms has a column labelled “Meaning and use of each word” with explanations on differences in meaning and usage.

- ③ *Ruigojiten* [類語辞典] (“*Dictionary of synonyms*”), Kodansha [講談社] 2008

Register labels included: “written language” [文章語 *bunshōgo*], “elegant, refined” [雅語 *gago*], “slang” [俗語 *zokugo*], “vulgarism” [卑語 *higo*], “baby talk” [幼児語・児童語 *yōjigo - jidōgo*], “female language” [女性語 *joseigo*], “male language” [男性語 *danseigo*].

Additional information is provided within definitions and in notes to single entries.

- ④ *Ruigigo tsukaiwake jiten* [類義語使い分け辞典] (“*Dictionary of synonym use*”), Kenkyūsha [研究社] 2007

No register labels. Entries contain explanations regarding registers.

As could be seen above, some thesauruses contain register labels, while others contain register information within single entries. Usage labels are used more often than in monolingual dictionaries, but their types are largely the same as in the monolingual dictionaries described in the previous section. Information on register is present not only in the form of labels, but also within additional explanations for single entries, and more detailed information can be found here than in monolingual dictionaries.

The same four pairs of adverbs listed in section 3. were analysed in these thesauruses, and the following information was found. Each piece of information is preceded by the number of the dictionary containing it.

**Table 6:** Register information for four pairs of adverbs in four different thesauruses

<i>tabun</i>	① — ② — ③ — ④ <i>Tabun</i> is a softer expression than <i>osoraku</i> .
<i>osoraku</i>	① — ② — ③ Has a rather formal nuance. ④ <i>Osoraku</i> is used in a formal polite style.
<i>zenzen</i>	① colloquial ② Comparing <i>zenzen</i> and <i>mattaku</i> , <i>zenzen</i> is a more informal expression. ③ — ④ —
<i>mattaku</i>	① — ② Comparing <i>zenzen</i> and <i>mattaku</i> , <i>zenzen</i> is a more informal expression. ③ — ④ —
<i>zenbu</i>	① Colloquial ② — ③ Used in a more spoken context than <i>subete</i> and in more concrete cases. ④ <i>Zenbu</i> means all parts and can be thought of as the spoken language variety of <i>subete</i> . However one part of it overlaps with the usage of <i>minna</i> , which is misleading.
<i>subete</i>	① colloquial ② — ③ — ④ <i>Zenbu</i> means all parts and can be thought of as the spoken language variety of <i>subete</i> . However one part of it overlaps with the usage of <i>minna</i> , which is misleading.

<i>ichiban</i>	<ul style="list-style-type: none"> <li>① colloquial</li> <li>② <i>Mottomo</i> is the more common word</li> <li>③ An expression often used in everyday conversation, with the meaning of <i>mottomo</i>.</li> <li>④ <i>Ichiban</i> is an expression used in spoken language, while <i>mottomo</i> is used in written language.</li> </ul>
<i>mottomo</i>	<ul style="list-style-type: none"> <li>① colloquialism</li> <li>② <i>Mottomo</i> is the more common word.</li> <li>③ An expression often used in everyday conversation and carries the meaning of <i>mottomo</i>.</li> <li>④ <i>Ichiban</i> is an expression used in spoken language, while <i>mottomo</i> is used in written language.</li> </ul>

When looking at the description of each word we see that register information is added to almost all words. Furthermore, such information is often given not only as a label, but in the form of an explanation. The thesauruses also differ from monolingual dictionaries in structure: synonyms referring to the same meaning are gathered in one entry, and there are therefore many comparisons between synonyms. The strength of this kind of description is the possibility of gaining information about the register of the searched word and its synonyms at the same time. This kind of description method would be welcome also in learner's dictionaries.

## 5. Information gained from corpora

In the preceding sections, register information was surveyed in learner's dictionaries, Japanese dictionaries and thesauruses, but it was found to be insufficient from the point of view of the situation in which a word is used. The following paragraphs discuss a method to describe such information in detail.

In recent years, corpora have been compiled for Japanese, and in 2011 the National Institute for Japanese Language and Linguistics constructed the Balanced Corpus of Contemporary Written Japanese (BCCWJ - *Gendai nihongo kakikotoba kinkō kōpasu* 現代日本語書き言葉均衡コーパス, Maekawa, 2008, 2011). The online corpus search application *chūnagon* has also been developed, making it possible to easily find collocation and other information by combining different search criteria. However, it is not possible to download the entire text with *chūnagon*<sup>5</sup>, making it difficult to analyse register information in detail. Although there is also a CD which contains the whole text, the absence of searching tools makes it difficult to analyse the text without a certain knowledge of natural language processing or programming.

<sup>5</sup> <https://chunagon.ninjal.ac.jp/>

In 2012 however, the online searching system NINJAL-LWP for BCCWJ<sup>6</sup> (hereafter abbreviated to NLB) was developed (Pardeshi, 2012). The main feature of this system is corpus search tool for lexical profiling which offers a comprehensive picture of the collocational and grammatical patterns of lexical words (nouns, verbs, adjectives and adverbs). The system also offers information regarding the frequency of each word in each of the subcorpora of BCCWJ (books, Diet minutes, Yahoo! Q&A (*Chiebukuro*), Yahoo! Blogs) in terms of tokens per million words. Within the books subcorpus, the system can also show separately the number of tokens per million words only within dialogues or only within prose. This information is useful for grasping the situations and contexts in which each word tends to be used.

The four pairs of adverbs were analysed using NLB to determine what information can actually be acquired about them using this corpus and tool.

**Table 7:** Frequency of occurrence of the four pairs of adverbs analysed  
(no. of tokens per million words)

	Books		Diet minutes	Yahoo! Q&A	Yahoo! Blogs
	Prose	Dialogues			
<i>tabun</i>	24.8	84.93	33.23	121.07	60.02
<i>osoraku</i>	84.52	89.41	112.39	68.97	44.6
<i>zenzen</i>	23.26	142.27	70.09	130.62	107.95
<i>mattaku</i>	231.99	289.19	217.12	266.31	200.48
<i>zenbu</i>	59.9	199.96	191.54	106.82	96.28
<i>subete</i>	376.14	299.05	164.75	224.52	229.65
<i>ichiban</i>	57.32	129.01	19.74	46.67	30.84
<i>mottomo</i>	218.31	84.39	77.54	42.73	80.02

Table 7 shows the tendency of occurrence for each word.

In the book corpus, the adverbs *tabun*, *zenzen*, *zenbu* and *ichiban* tend to occur frequently in the dialogues, while the remaining prose contains more occurrences of *subete* and *mottomo*. *Osoraku* and *mattaku* were found more or less equally often in both types of texts. In the diet minutes, which record questions and answers in a very formal setting, the following tendency can be seen: *tabun* < *osoraku*, *zenzen* < *mattaku*, *zenbu* > *subete* and *ichiban* < *mottomo*. The texts of Yahoo!Q&A and Yahoo!Blogs are written by ordinary people, without any editing or monitoring. Yahoo!Q&A (*Yahoo!Chiebukuro*) consists of questions on any topic written by anonymous users and answers to these questions by other anonymous users; it can therefore be expected that these texts are written with some consideration and respect for the expected readers. The Yahoo!Blog subcorpus consists of texts in which writers

<sup>6</sup> <http://ninjal-lwp-bccwj.ninjal.ac.jp/>

often express their own opinions. Both the Yahoo!Q&A and the Yahoo!Blogs subcorpus contain texts of very different styles, and it is therefore difficult to draw any conclusion from the numbers in Table 7. It is however apparent that *tabun*, *mata* and *subete* tend to occur frequently.

On the basis of the above results, the following observations can be made about the tendencies of use of the four pairs of adverbs.

- *Tabun* occurs in dialogue-like texts. It also tends to occur in subjective texts.
- *Osoraku* occurs in explanatory texts. A tendency of usage in informal spoken settings can be observed.
- *Zenzen* occurs in dialogue-like texts.
- *Mattaku* occurs in dialogue-like and explanatory texts. However, it also tends to be used in formal spoken language and when the writer is very much aware of the readers of a text.
- *Zenbu* occurs in dialogue-like contexts.
- *Subete* tends to occur in explanatory texts and when the writer is very much aware of the readers of a text.
- *Ichiban* occurs in dialogue-like contexts.
- *Mottomo* occurs in explanatory contexts. It also tends to be used in formal spoken settings.

The texts gathered in these four corpuses are not necessarily homogeneous. It is therefore only possible to observe certain tendencies of each word's occurrence. However, by surveying these tendencies and frequencies, and further analysing register differences as they are reflected in patterns and tendencies of word usage in BCCWJ, it should be possible to describe the register of single words.

## 6. Conclusion

The present paper presented an overview of register labelling in learner's and monolingual dictionaries and thesauruses.

Learners' dictionaries contain relatively plentiful register information from the viewpoint of interpersonal relationships. However, the following kinds of information are still insufficiently provided: (1) information on how different synonyms are used differently in writing according to different registers and (2) register labels such as "archaic" an/or "literary" for words that are slightly outside everyday use but within target vocabulary for middle or advanced learners of Japanese.

On the other hand, there are more types of register labels in Japanese monolingual dictionaries for native speakers than in learners' dictionaries, but register labels tends to be assigned mostly to words with a clearly limited usage, such as "archaic", "elegant (poetic)", "slang", and "(technical) jargon", while there is hardly any information on register regarding everyday words. Also, register labels provided are often based on Miyajima's three-tier distinction of "everyday words" in the middle, with "written words" and "colloquial words" on both ends, without much regard to interpersonal aspects and situation of use, which would be more informative for learners of Japanese.

Thesauruses contain more register information than monolingual Japanese dictionaries, which is especially useful when comparing synonyms, to know which synonym is to be used in which context. Such discussions of the differences between synonyms of similar meaning would be welcome also in dictionaries for learners of Japanese as a second language.

Considering register information as has been found in the surveys described above, the following five groups of register labels can be considered necessary in dictionaries for learners of Japanese as a foreign language:

- 1) labels related to interpersonal relationships, for words which need to be used with careful consideration of the listener or reader, such as expressions indicating the speaker's respect for the listener or referent, or words which can be unpleasant to the listener;
- 2) labels for words used in spoken conversation, related to the setting in which the word is used, such as formal or informal occasions;
- 3) labels for words used mainly in writing, related to different settings, such as letters or other texts requiring formal formats, or texts such as reports which need to be written objectively;
- 4) labels for archaic expressions, rare words, literary and other expressions with limited use in contemporary language;
- 5) labels related to terminology from specialised fields.

By further dividing these five groups of labels and adding such detailed information to dictionary entries, it would be clear which words are used in what circumstances. With the help of such detailed register information, it would also be possible to clearly describe the use of synonyms, when including them in an entry. In addition, examples showing typical settings in which each word is used would also be of great benefit to learners.

In order to assign such detailed labels to dictionary entries, it is necessary to know in what contexts each word is used. This can be gleaned from corpora. In this paper, NLB was used to observe tendencies of use of some adverbs in different subcorpora. For a comprehensive and detailed lexicographic description, other specific and



homogeneous corpora are needed, such as spoken corpora or corpora of academic writing.

## 7. Further research

With the creation of large-scale corpora for Japanese, the resources needed for the analysis of lexical meaning, use, and collocations are gradually taking shape. However, in order to analyse lexical register, homogeneous specialised corpora need to be constructed, and this, in turn, requires the analysis and categorisation of textual characteristics. Such an analysis of different types of texts, aimed at further refining the labelling types mentioned above, will be the subject of further research.

## References

- Maebo, K. [前坊香菜子] (2009). Go no buntaiteki tokuchō ni kansuru jōhō ni tsuite no ichi kōsatsu – Kokugo jiten to ruigo jiten no chōsa kara [語の文体的特徴に関する情報についての一考察—国語辞典と類語辞典の調査から—]. *Hitotsubashi nihongo kyōiku kenkyū hōkoku* [一橋日本語教育研究報告]. 3: 50-60.
- Maekawa, K. (2008). Balanced Corpus of Contemporary Written Japanese. In Huang, C.-R., Mikami, Y., Hasida, K., & Tokunaga, T. (Eds.) *Proceedings of The 6th Workshop on Asian Language Resources (ALR 6)*, 101-102. Retrieved July 2012 from <http://www.aclweb.org/anthology-new/I/I08/I08-7.pdf#page=109>.
- Maekawa, K. (2011). Development of Japanese Corpora at the National Institute for Japanese Language and Linguistics: With Emphasis on Five Sources of Difficulty in Japanese Corpus Development. *Lexicography: Theoretical and Practical Perspectives (ASIALEX2011 Proceedings)*, 17-26.
- Miyajima, T. [宮島達夫] (1977). Tango no buntaiteki tokuchō [単語の文体的特徴]. In: *Muramatsu Akira kyōju kanreki kinen Kokugogaku to kokugoshi* [村松明教授還暦記念国語学と国語史]. Tōkyō: Meiji shoin [明治書院]. 871-903.
- Nakamichi, M. [中道真木男] (1989). Go no buntaiteki tokuchō [語の文体的特徴]. In: *Keesu sutadei Nihongo no goi* [ケーススタディ 日本語の語彙]. Tōkyō: Ōfūsha [桜楓社].
- Nihongo-kyōiku Gakkai [日本語教育学会編] (1987). *Nihongo-kyōiku jiten* [日本語教育事典]. Kihon-go, Kiso-go [基本語、基礎語]. 295. Taishukan shoten [大修館書店]
- Pardeshi, P. (2012). Compilation of Japanese Basic Verb Usage Handbook for JFL Learners: A Project Report. *Acta Linguistica Asiatica*, 2(2), 37-63.
- Yonekawa, A. [米川明彦] (2002). Dai 6 shō Isō to isōgo [第6章 位相と位相語]. In *Asakura Nihongo kōza 4 – Goi – imi* [朝倉日本語講座4 語彙・意味]. Tōkyō: Asakura shoten [朝倉書店].

## Dictionaries

- Benesse shinshū kokugo jiten* [ベネッセ新修国語辞典] (2nd Ed.). (2012). Tokyo: Benesse [ベネッセ].
- Gakken gendai hyōjun kokugo jiten* [学研現代標準国語辞典] (2nd Ed.). (2011). Tokyo: Gakken Kyōiku Shuppan [学研教育出版].
- Informative Japanese Dictionary - Nihongo o manabu hito no jiten [日本語を学ぶ人の辞典]. (1995). Tokyo: Shinchosha [新潮社].
- Kadokawa ruigo shin jiten* [角川類語新辞典]. (2002). Tokyo: Kadokawa [角川].
- Kenkyusha's English-Japanese Japanese-English Learners' Pocket Japanese-English Learners' Dictionary – Kenkyūsha einichi nichiei poketto jiten [研究社 英日・日英ポケット辞典]. (1992). Tokyo: Kenkyusha [研究社].
- Kodansha's Furigana Japanese Dictionary - Furigana eiwa-waei jiten [ふりがな和英・英和辞典]. (1999). Tokyo: Kodansha [講談社].
- Tuttle Concise Japanese Dictionary - Tuttle konsaisu eiwa-waei jiten [タトル・コンサイス英和・和英辞典] (2008). Rutland, Vermont, & Tokyo: Tuttle.
- Reikai shin kokugo jiten* [例解新国語辞典] (8th Ed.). (2012). Tokyo: Sanseido [三省堂].
- Ruigojiten* [類語辞典]. (2008). Tokyo: Kōdansha [講談社].
- Tsukaikata no wakaru ruigo reikai jiten shinsouhan* [使い方のわかる類語例解辞典新装版]. (2003). Tokyo: Shōgakukan [小学館].