Boot-strapping a WordNet using multiple existing WordNets

Francis Bond, Hitoshi Isahara, Kyoko Kanzaki, Kiyotaka Uchimoto

Language Infrastructure Group
National Institute of Information and Communications Technology (Japan)
bond@ieee.org,{bond,isahara,kanzaki,uchimoto}@nict.go.jp

Abstract

In this paper we describe the construction of an illustrated Japanese Wordnet. We bootstrap the Wordnet using existing multiple existing wordnets in order to deal with the ambiguity inherent in translation. We illustrate it with pictures from the Open Clip Art Library.

1. Introduction

It is rare for languages to have many freely available lexical semantic resources. In particular, few languages have as many as English. This lack of resources slows down both theoretical and applied research into language meaning across languages.

Consider the case for Japanese. There are at least two excellent thesauruses available: the Japanese Synonym Dictionary (Hamanishi and Ono, 1990) and Goi-Taikei, a Japanese Lexicon (Ikehara et al., 1997). Unfortunately, although they were both used extensively in machine translation research, the published resources focus almost entirely on Japanese. Further, the thesauruses are proprietary, which makes it extremely hard to share and build on the results of research using them.

To alleviate this problem, we are investigating methods to quickly and efficiently build a Japanese version of Wordnet (Fellbaum and Vossen, 2007). This wordnet is based on the structure of the English wordnet — Japanese near synonyms are added to the existing English synsets. For example, the English synset consisting of `seal#n#9` "any of numerous marine mammals that come on shore to breed; chiefly of cold regions"[1] has the following Japanese words associated with it: アザラシ *azarashi* "seal" and 海豹 *azarashi* "seal".

The WordNet project at Princeton has been a resounding success, creating a resource that is widely used in research (Fellbaum, 1998) and emulated in many languages (Vossen, 1998). In order for a lexical resource to be widely adopted it must be both **accesible** and **usable**. The Princeton WordNet is accessible due to its being released under a non-restrictive license; and usable because it has not just precise information but also reasonable coverage, especially of common words.

Because of this success, there have been many projects to build wordnets for other languages. One of the first was the EuroWordNet project, which built wordnets for several European languages (Vossen, 1998). Unfortunately, most of the wordnets are neither as accesible as the Princeton WordNet, due to more restrictive licenses, nor as usable due to more limited cover. Recently, the Global WordNet grid has tried to add even more languages, making the data as accesible as possible (Fellbaum and Vossen, 2007).

There have been several initiatives to create a Japanese wordnet, but none of them have yet produced something that is both accessible and usable. Hayashi (1999) created a translation of the entire noun part of the Princeton WordNet, including both synsets and glosses. This produced a very usable resource, but it was unfortunately not at all accessible. Koide et al. (2006) looked at combining EDR (EDR, 1990) with Princeton WordNet, but did not get beyond converting them both to RDF representations. Kaji and Watanabe (2006) presented a method of translating synsets from English to Japanese using corpus based contexts to improve accuracy, but only tested this on a few words. More recently, Cook (2008) produced a Multi-Lingual Semantic Network by translating monosemous parts of the Princeton WordNet into Japanese, Chinese and German. He also made an interface for browsing and amending the network. This data is accessible, as it is released under an open license, but loses a little on usability as most monosemous entries are for less frequent words.

The amount of previous work shows the great interest and value of producing a Japanese WordNet. We therefore decided to construct one as follows: First, automatically translate the Princeton WordNet into Japanese. Second, manually check the most frequent 20,000 synsets. Third, link the synsets to a corpus. Fourth, release the data under an open license. As we said earlier, this WordNet is based on the structure of the English wordnet: Japanese near synonyms are added to the existing English synsets. Adapting it more fully to Japanese is left to future research. More details of the overall project are given in Isahara et al. (2008).

The obvious way to do add Japanese to the English WordNet is by translating the entries using an English-Japanese dictionary. The problem with this is that bilingual dictionaries are not marked with WordNet senses, if we look up *seal* we get over 30 entries, including 判子 *seal* "stamp" and 海軍特殊部隊 *gaiguntokushubutai* "Navy Seal". We need to associate these candidates with the appropriate WordNet senses.

Our method takes advantage of the existence of wordnets in multiple languages, and uses them to sense disambiguate the translations. We were able to build it quickly and efficiently using the results of existing work on building wordnets and lexicons, and we intend to make it freely available so that other people can build on it.

2. Lexical Resources

In this section we describe the resources we have used. Most of them are open resources.

[1] All examples are from WordNet 3.0.

Part of Speech	Number of Synsets English	French	Spanish	German
Noun	82,115	17,826	7,902	9,951
Verb	13,767	4,919	3,775	5,166
Adjective	18,156	0	3,879	15
Adverb	3,621	0	0	0
Total	117,659	22,745	15,556	15,132

Table 1: Sizes of the Wordnets used

Part of Speech	Number of Synsets $s > 10$	$s > 1$	All
Noun	9,243	36,432	42,725
Verb	2,991	9717	10,321
Adjective	629	6,283	8,915
Adverb	9	1,317	1,726
Total	12,872	53,749	63,687

Table 3: Japanese Synsets by score

2.1. Wordnets

We use four wordnets, summarized in Table 1. The largest is the English Wordnet v3.0 (Fellbaum, 1998) with 117,659 entries. The EuroWordnets are considerably smaller, ranging from 15,132 for German up to 22,745 for French (Vossen, 1998), consisting mainly of nouns with some verbs. All of them share the same structure — a collection of synsets joined to make a semantic network.

Because Wordnet keeps growing, both in size and complexity synsets can split up or even potentially merge across versions. The data for German was based on 1.5 and French and Spanish on 1.6. We mapped them into 3.0 using the mappings from Daude et al. (2003). When a synset mapped to more than one synset, we simply linked it to the most highly weighted one.

2.2. Lexicons

We use JMDict, the Japanese→Multilingual dictionary created by Jim Breen (Breen, 2004) for Japanese-English/French/German. We did not use its proper name dictionary, as wordnet does not have a lot of names. JMDict is widely used, and is increasing in size at the rate of almost 1,000 entries a month (Bond and Breen, 2007). To supplement this we also used the EDR Japanese-English lexicon (`http://www2.nict.go.jp/r/r312/EDR/index.html`) and the last downloadable version of the Japanese-English Life Science Dictionary Project (v4) (`http://lsd.pharm.kyoto-u.ac.jp/ja/index.html`). For Japanese-Spanish, we used a small dictionary downloaded from `http://aulex.ohui.net/` (Goihata) and licensed under the GPL. The sizes of these lexicons are listed in Table 2.

The lexical resources, are, as always, not evenly distributed amongst the world's languages — Japanese-English has the most resources, followed by German, then French and then Spanish.

3. Creating the Japanese Wordnet

The approach we are taking to build the Japanese Wordnet is the standard **expand approach**: "translate WordNet synsets to another language and take over the structure" (Vossen, 2005). We did this both to keep a compatible structure with WordNet, and because we had access to a variety of resources to make the task easier. Our main innovation is that we are using WordNets in multiple languages to disambiguate the Japanese translations, thus providing more reliable estimates.

Consider the following two synsets for *bat*, with their translation shown in Figure 1:

 bat#n#1, chiropteran (nocturnal mouselike mammal with forelimbs modified to form membranous wings ...)

 bat#n#5 (a club used for hitting a ball in various games)

The Japanese-English lexicon has two translations for *bat* 蝙蝠 *koumori* "bat (mammal)" and バット *batto* "bat (club)". However, because there is no way of distinguishing between them we get a mixture of the meanings with 蝙蝠 *koumori* "bat#n#1" and バット *batto* "bat#n#5". *chiropteran* is not in any of the JE lexicons, and `bat#n#5` has no synonyms. Therefore using only the English Wordnet as source and Japanese⇔English lexicons there is no way to disambiguate them.

However, both synsets are also in the French wordnet: `bat#n#1` is *chauve-souris* and `bat#n#5` is *batte, gourdin*. These are not ambiguous in the same way: *chauve-souris* goes only to *koumori* and *batte* only to *batto*. Thus, if we can match through two languages, the mapping is much more likely to be the correct sense. The links are shown in Figure 1.

Similar approaches have been used to make new bilingual dictionaries: for example, linking Japanese-Malay through Japanese-English, English-Malay, Japanese-Chinese and Chinese-Malay (Bond et al., 2001). The difference here is that the original linking is done through the Wordnet synsets: we are effectively trying to translate a super-synset with synonyms in up to four languages (En, De, Fr, Es).

The actual algorithm we used was as follows:

- For each synset in WordNet 3.0
  - Find equivalents in WN-{Fr,Es,De}
  - Look up translations for all equivalents $\{J_e\}$, $\{J_f\}$, $\{J_s\}$, $\{J_d\}$
  - Rank Japanese equivalents
    score $s = |\text{links}| + 10$ for links in two languages

The result is a wordnet with multiple Japanese candidates for most synsets, with a confidence score s equal to the number of bilingual links plus a ten-point bonus for being linked in multiple languages.

4. Results and Evaluation

In this section we report on how many synsets we could translate into Japanese, and with what confidence.

The results are summarized in Table 3. We have found some kind of translation for 63,687 out of the possible 117,007 synsets in Wordnet 3.0 (54.4%). Of these, the EuroWordnet data played a role in over 15,000 synsets. 12,872

Part of Speech	Number of Word-Pairs					
	ja-en			ja-de	ja-fr	ja-es
	JMDict	EDR	Lifsci	JMDict	JMDict	Goihata
Noun	165,984	504,450	44,567	143,753	24,348	0
Verb	22,209	184,250	4,741	26,502	7,762	133
Adjective	16,861	44,961	11,212	17,121	4,582	70
Adverb	6,180	20,125	1,266	5,915	1,478	0
Unknown	3	0	0	0	0	3,548
Total	225,803	758,568	62,210	199,260	39,447	3,751

Table 2: Size and Distribution of the various Lexicons

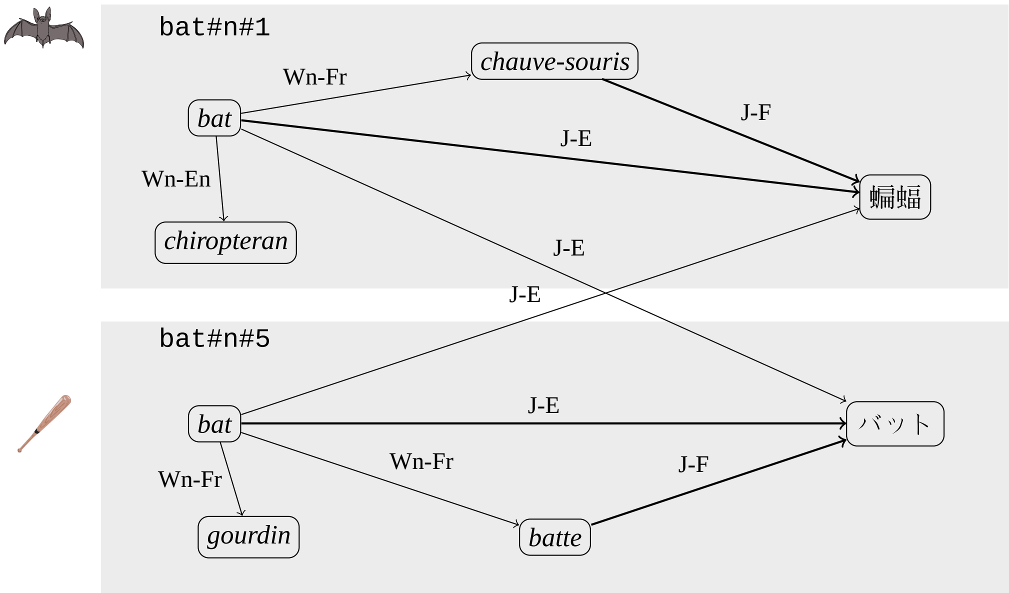

Figure 1: Linking with Multiple Wordnets

Part of Speech	Number of Synsets		
	$s > 10$	$s > 1$	All
Noun	2,429	3,264	3,279
Verb	656	988	993
Adjective	153	586	653
Adverb	0	0	0
Total	3,238	4,838	4,925

Table 4: Base Japanese Synsets by Score (s) for the Base Synsets

synsets had at least one translation candidate confirmed in two or more languages, and 53,749 were confirmed in multiple lexicons.

The results restricted to the 5,000 common base synsets which occupy central positions in the wordnet structures (Fellbaum and Vossen, 2007) are given in Table 4. In this case our cover is almost complete (4,925/5,000 = 98.5%). Most of the entries in Euro WordNet are from these base synsets, and the majority of our translations (64.6%) match in two or more languages. Our coverage is excellent for the base synsets, and good overall: larger than any of the existing non-English WordNets.

To test precision, we evaluate translation candidates by judging the suitability of all of the base synset translation candidates (this is actually part of preparing the WordNet

	Appropriate Translation Candidates			
	$s > 10$	$10 > s > 1$	$s = 1$	All
Base	55.30%	39.64%	21.25%	26.56%

Table 5: Base Noun Candidate Precision

for release). The results are given in Table 5. Translations matching in multiple languages are markedly better than those matching in a single language (55.30% vs 21.25%), matching in multiple dictionaries in a single language is also an indicator of higher accuracy.

For both evaluation methods, translating synsets using multiple languages gives higher precision and lower recall.

5. Illustrating WordNet

In order to make the sense distinctions more visible we also semi-automatically link synsets to illustrations from the Open Clip Art Library (OCAL: Phillips (2005)). This adds a new modality to the knowledge linked in the semantic net. Illustrations of concepts are useful for a variety of tasks. One is pedagogical — it is useful to have pictures in learners' dictionaries. Another is in cross-cultural communication - for example in Pangea, where children use pictons (small concept representing pictures) to write messages (Takasaki and Mori, 2007).

We use the collection of OCAL images distributed as SVG (scalable vector graphic) images in the Ubuntu Fiesty distribution based on the OCAL release of October 2005 (v 0.18). It contains 8,107 images (with some duplicates), organized in a shallow file hierarchy. Currently, some 4,000 new images have been added to the OCAL, but we have not yet processed them.

Each image is associated with a collection of explicit metadata, including a title, description and a set of tags, all of which are recommended rather than obligatory. SVG images are written in XML, the metadata is embedded within using the Creative Commons' metadata standard.

We take advantage of the metadata associated with each image to associate the image to a specific synset. The basic idea is to look for metadata associated with a word and its hypernym: if we can find a match of this combination in Wordnet, then we consider it a valid illustration for that synset. For example, for `bat_orlando_karam_.svg`, its title is *bat* and it is tagged as *mammal*. We look in wordnet for hypernym synsets of *bat* that include *mammal* and find the following: `bat#n#1` $\subset$ `placental#n#1` $\subset$ `mammal#n#1`. Therefore, this picture illustrates `bat#n#1` rather than the other synsets associated with the word *bat*

There are several sources of metadata. We first use explicit metadata such as TITLE for the root word and TAGS for the hypernyms. If there is no explicit metadata (which is true for around a third of the images) or we couldn't find a match, then we look for implicit metadata. We take the basename `bat_orlando_karam` and delete any numbers from the end. We also add directory names to the tag list (`animals/mammals/bat_orlando_karam_.svg`): in this case "mammal" and "animal". Finally, we match the tags against each other.[2]

Using the tags allows us to largely solve the problem of **Image Sense Disambiguation** for those pictures we identify. However, it does not solve all of the problems raised in (Alm et al., 2006), in particular the problems of **depiction** (is a sign with a train on it an example of *train*?) and **partial display** (is a picture of a dog's head a good illustration for *dog* or *head* or neither or both?).

There are 956 illustrations which match, illustrating 758 synsets. All the successful links were of nouns. Most matches are of concrete objects, and generally of the base synsets. The synset with the most matches is `smiley#n#1` "an emoticon of a smiling face" which has 33 illustrations.

We have only linked a small subset of illustrations (936 out of 8,107 images) and an even smaller proportion of wordnet (758 out of 82,115 noun synsets). However, these figures are better than they seem — many of the illustrations were not suitable in the first place. And any illustrated synset also (in theory) illustrates its hypernyms, so we have indirectly illustrated far more than 758 synsets.

[2] We do some normalization when we look words up in wordnet: if we can't find a word as is, we then look it up downcased, without spaces, and in singular form: *paint brushes* $\rightarrow$ {*paintbrushes*, *paint brush*, *paintbrush*}.

6. Discussion

In this paper we presented a method of automatically producing a Japanese WordNet of reasonable quantity by cross checking senses across different languages.

In future work we will manually check the most frequent synsets, sense annotate a small corpus and release the data. We hope it will then be used along with other projects such as the Global Wordnet Grid (Fellbaum and Vossen, 2007) and the Multi-lingual Semantic Network (Cook, 2008) to produce an even more useful resource.

Our results confirm one of the advantages of the global wordnet grid: available high quality existing resources makes it easier to build more. Now we need to make these resources as easily available as possible to enable even further progress. Of course, not all languages have as many available resources as Japanese. However, bilingual lexicons are much more common than wordnets, so it makes sense to use the more common lexical resources to bootstrap the rarer one.

An example of the complete results for the synset `tree#n#1` "a tall perennial woody plant ..." is given in Figure 2. Each synset is given a name consisting of the highest scoring Japanese match with the first English entry for that synset, in this case "木/tree". This is the format we use for browsing the results of our matching.

The Japanese matches are shown in three sets: $s > 10$ (◎); $10 > s > 1$ (○); $s = 1$ (●). In this case 2/3 of the top set (◎) are good matches; the third ツリー*tsurii* "tree" is mainly used for Christmas trees. The middle set (○) consists soley of 木本 *mokuhon* "woody plant", which is the immediate hypernym of `tree#n#1`. The second member of the last set is also good, the rest are irrelevant. The matches for the hyponyms (hypo) are mainly good: the lower down the hierarchy the less likely words are to be ambiguous.

We have also made a first step towards illustrating wordnet. We expect the number of linked illustrations to grow due to the following factors (a) more images (and better tags) being added to the OCAL; (b) more words being added to wordnet and (c) improvements in the matching algorithm. In the long term, we would like to integrate the wordnet linking into the Open Clip Art Language workflow, so that new images can be tagged as they are added to the library. We hope that the link to definitions, examples and multilingual equivalents will provide even more motivation to artists to add accurate and detailed meta-tags.

From the point of view of the open clip art project, tagging illustrations with wordnet synsets will allow people to search for pictures more effectively. In particular, they can associate the image with its synonyms and hypernyms — someone looking for *pinniped* or *aquatic mammal* could find the image associated with *seal*. In addition, thanks to the global wordnet grid, we can do this in multiple languages: `bat#n#1` is linked to *chauve-souris* "bat" in the French wordnet, *Fledermaus* "bat" in the German wordnet, 蝙蝠 *koumori* "bat" in the Japanese wordnet and so on.

7. Further Work

In the immediate future we plan to hand correct more entries, and sense tag a small corpus. We plan to release

13104059-n --- (木/tree) ‡

a tall perennial woody plant having a main trunk and branches forming a distinct elevated crown; includes both gymnosperms and angiosperms

◎ 木, 樹, ツリー,
○ 木本,
● あずま屋, 樹木, 高木, 東屋, 喬木, 木物, 天然木, 四阿, 成木, シャフト,

〔noun.plant - 107〕

syns: *tree#n#1*

hype: 木本/ligneous_plant

hypo: ？/yellowwood_tree ？/lancewood ？/negro_pepper ？/anise_tree ？/drimys_winteri ？/zebrawood ？/granadilla_tree アカシア/acacia 紫檀/red_sandalwood ？/albizia ？/elephant's_ear ？/inga ？/inga_edulis ？/inga_laurina ギンネム/white_popinac ？/wild_tamarind ？/nitta_tree ？/huamachil ？/dita_bark ？/conessi ？/meryta_sinclairii 距爪/cockspur 蛸の木科/screw_pine ？/hoheria_populnea ？/plagianthus_betulinus ？/tulipwood_tree ？/bombax_malabarica ？/montezuma ？/pseudobombax_ellipticum ？/elaeocarpus_grandis ？/jamaican_cherry ？/break-axe ？/bottle-tree 梧桐/phoenix_tree ？/maple-leaved_bayur ？/tarrietia_argyrodendron サンバ/arere ライム/basswood ？/silver_tree 山椒/orites_excelsa ？/stenocarpus_sinuatus ？/beefwood ？/casuarina ブナ科/beech クリ/chestnut_tree ？/oak_chestnut ？/giant_chinkapin ？/lithocarpus_densiflorus ？/southern_beech オーク/oak 樺/birch ハンノキ/alder_tree 赤四手/hornbeam ？/hop_hornbeam ？/fringe_tree 灰分/ash_tree 木犀/osmanthus_americanus ？/dhawa ？/conocarpus_erectus ？/white_mangrove 楊梅/jamaica_bayberry 樹脂状物質/pterocarpus_marsupium 紫檀/red_sandalwood ？/sabinea_carinalis ？/scarlet_wisteria_tree 槐/chinese_scholar_tree ...

msub: 白身/sapwood 心材/duramen

mprt: 株/tree_stump 頂/crown 肢/limb 幹/tree_trunk 節/burl

hmem: 森/woods

es: árbol

fr: arbre

Search for Japanese word in Synsets

Japanese Word: 言葉 Search WN

Figure 2: Example Entry for 木/tree

the manually checked subset of WordNet sometime in June 2008. Rather than have a single maintainer and major releases, we hope to maintain the WordNet as a community resource, along the lines suggested by Charoenporn et al. (2008). Here a wiki-like tool is used to allow people to extend and amend the WordNet, with final changes checked by moderators. In this release, we also intend to add high confidence automatic entries (unambiguous translations of monosemous words) as suggested by Cook (2008) and Charoenporn et al. (2008). This should add another 12,000 or so entries to the hand checked subset.

Currently, we have not made any new synsets for Japanese words whose meaning does not cleanly map to an English synset, although we know that these are necessary. As well as extending these synstes we will also extend the information about lexical relations to include those not in the original Princeton WordNet, such as attribute-instance relations (Kanzaki et al., 2008) and explicit linking of orthographic variants: *color/colour* or 蝙蝠/コウモリ *koumori* "bat".

In other major extensions we are hoping to do the following:

- Link the wordnet to other ontologies, such as concepts in EDR (EDR, 1990), GoiTaikei (Ikehara et al., 1997) and CoreNet (Choi and Bae, 2003).
- Sense tag a variety of corpora.
- Use the WordNet data as a module for Japanese in the METEOR Automatic Machine Translation Evaluation System (Lavie and Agarwa, 2007).

8. Conclusions

In this paper we described the construction of the illustrated Japanese Wordnet. We bootstrapped the Wordnet using existing wordnets and bilingual lexicons. We were able to produce Japanese translations for 98% of the core classes, over half of them with high confidence.

9. References

Cecilia Ovesdotter Alm, Nicolas Loeff, and David A. Forsyth. 2006. Challenges for annotating images for sense disambiguation. In *Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora 2006*, pages 1–4, Sydney, Australia, July. Association for Computational Linguistics.

Francis Bond and James Breen. 2007. Semi-automatic refinement of the JMdict/EDICT Japanese-English dictionary. In *13th Annual Meeting of The Association for Natural Language Processing*, pages 364–367, Kyoto.

Francis Bond, Ruhaida Binti Sulong, Takefumi Yamazaki, and Kentaro Ogura. 2001. Design and construction of a machine-tractable Japanese-Malay dictionary. In *MT Summit VIII*, pages 53–58, Santiago de Compostela, Spain.

J. W. Breen. 2004. JMDict: a Japanese-multilingual dictionary. In *Coling 2004 Workshop on Multilingual Linguistic Resources*, pages 71–78, Geneva.

Thatsanee Charoenporn, Virach Sornlerlamvanich, Chumpol Mokarat, and Hitoshi Isahara. 2008. Semi-automatic compilation of Asian WordNet. In *14th Annual Meeting of the Association for Natural Language Processing*, pages 1041–1044, Tokyo.

Key-Sun Choi and Hee-Sook Bae. 2003. A Korean-Japanese-Chinese aligned wordnet with shared semantic hierarchy. In *Proceedings of the International Conference on Natural Language Processing and Knowledge Engineering*, pages 91–96.

Darren Cook. 2008. MLSN: A multi-lingual semantic network. In *14th Annual Meeting of the Association for Natural Language Processing*, pages 1136–1139, Tokyo.

Jordi Daude, Lluis Padro, and German Rigau. 2003. Validation and tuning of wordnet mapping techniques. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP'03)*, Borovets, Bulgaria.

EDR. 1990. Concept dictionary. Technical report, Japan Electronic Dictionary Research Institute, Ltd, April.

Christiane Fellbaum and Piek Vossen. 2007. Connecting the universal to the specific: Towards the global grid. In *First International Workshop on Intercultural Collaboration (IWIC-2007)*, pages 2–16, Kyoto.

Christine Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Masando Hamanishi and Susumu Ono, editors. 1990. *Ruigo Kokugo Jiten [Japanese Synonym Dictionary]*. Kadokawa Shoten, Tokyo. (in Japanese).

Yoshihiko Hayashi. 1999. Translating WordNet noun part into Japanese for cross-language natural language applications. In *Technical Reports of SIG on Natural Language Processing NL130-10*, pages 73–80. (in Japanese).

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Ooyama, and Yoshihiko Hayashi. 1997. *Goi-Taikei — A Japanese Lexicon*. Iwanami Shoten, Tokyo. 5 volumes/CDROM.

Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama, and Kyoko Kanzaki. 2008. Development of the Japanese WordNet. In *Sixth international conference on Language Resources and Evaluation (LREC 2008)*, Marrakech. (this volume).

Hiroyuki Kaji and Mariko Watanabe. 2006. Automatic construction of Japanese WordNet. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, May.

Kyoko Kanzaki, Francis Bond, Noriko Tomuro, and Hitoshi Isahara. 2008. Extraction of attribute concepts from Japanese adjectives. In *Sixth international conference on Language Resources and Evaluation (LREC 2008)*, Marrakech. (this volume).

Seiji Koide, Takeshi Morita, Takahira Yamaguchi, Hendry Muljadi, and Hideaki Takeda. 2006. OWL expressions on WordNet and EDR. In *AI society Semantic Web Ontology SIG 13*, SIG-SWO-A601-03. (in Japanese).

Anton Lavie and Abhaya Agarwa. 2007. METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In *Proceedings of Workshop on Statistical Machine Translation at ACL-2007*, Prague.

Jonathan Phillips. 2005. Introduction to the open clip art library. `http://rejon.org/media/writings/ocalintro/ocal_intro_phillips.html`. (accessed 2007-11-01).

Toshiyuki Takasaki and Yumiko Mori. 2007. Design and development of a pictogram communication system for children around the world. In *First International Workshop on Intercultural Collaboration (IWIC-2007)*, pages 144–157, Kyoto.

Piek Vossen, editor. 1998. *Euro WordNet*. Kluwer.

Piek Vossen. 2005. Building wordnets. `http://www.globalwordnet.org/gwa/BuildingWordnets.ppt`.