Robust Ontology Acquisition from Machine-Readable Dictionaries

Eric Nichols
Nara Inst. of Science and Technology
Nara, Japan
`eric-n@is.naist.jp`

Francis Bond
NTT Communication Science Labs
Nippon Telegraph and Telephone Co.
Keihanna, Japan
`bond@cslab.kecl.ntt.co.jp`

Daniel Flickinger
CSLI
Stanford University
California, U.S.A.
`danf@csli.stanford.edu`

Abstract

In this paper, we outline the development of a system that automatically constructs ontologies by extracting knowledge from dictionary definition sentences using Robust Minimal Recursion Semantics (RMRS), a semantic formalism that permits underspecification. We show that by combining deep and shallow parsing resources through the common formalism of RMRS, we can extract ontological relations in greater quality and quantity. Our approach also has the advantages of requiring a very small amount of rules and being easily adaptable to any language with RMRS resources.

1 Introduction

Ontologies are an important resource in natural language processing. They have been shown to be useful in tasks such machine translation, question answering, and word-sense disambiguation, among others where information about the relationship and similarity of words can be exploited. While there are large, hand-crafted ontologies available for several languages, such as WordNet for English [Fellbaum, 1998] and GoiTaikei for Japanese [Ikehara *et al.*, 1997], these resources are difficult to construct and maintain entirely by hand. They are, however, of proven utility in many NLP tasks, such as PP-attachment, where results using WordNet approach human accuracy (88.1% vs 88.2%), while the best methods using automatically constructed hierarchies still lag behind (at 84.6%) [Pantel and Lin, 2000]. Therefore, there is still a need to improve methods of both fully-automated and supervised construction of ontologies.

There is a great deal of work on the creation of ontologies from machine readable dictionaries (a good summary is [Wilkes *et al.*, 1996]), mainly for English. Recently, there has also been interest in Japanese [Tsurumaru *et al.*, 1991; Tokunaga *et al.*, 2001; Bond *et al.*, 2004]. Most approaches use either a specialized parser or a set of regular expressions tuned to a particular dictionary, often with hundreds of rules. In this paper, we take advantage of recent advances in both deep parsing and semantic representation to combine general purpose deep and shallow parsing technologies with a simple special relation extractor.

Our basic approach is to parse dictionary definition sentences with multiple shallow and deep processors, generating semantic representations of varying specificity. The semantic representation used is robust minimal recursion semantics (RMRS: Section 2.2). We then extract ontological relations using the most informative semantic representation for each definition sentence.

In this paper we discuss the construction of an ontology for Japanese using the the Japanese Semantic Database Lexeed [Kasahara *et al.*, 2004]. The deep parser uses the Japanese Grammar JACY [Siegel and Bender, 2002] and the shallow parser is based on the morphological analyzer ChaSen.

We carried out two evaluations. The first gives an automatically obtainable measure by comparing the extracted ontological relations by verifying the existence of the relations in exisiting WordNet [Fellbaum, 1998]and GoiTaikei [Ikehara *et al.*, 1997] ontologies. The second is a small scale human evaluation of the results.

2 Resources

2.1 The Lexeed Semantic Database of Japanese

The Lexeed Semantic Database of Japanese is a machine readable dictionary that covers the most common words in Japanese [Kasahara *et al.*, 2004]. It is built based on a series of psycholinguistic experiments where words from two existing machine-readable dictionaries were presented to multiple subjects who ranked them on a familiarity scale from one to seven, with seven being the most familiar [Amano and Kondo, 1999]. Lexeed consists of all open class words with a familiarity greater than or equal to five. An example entry for the word ドライバー *doraibā* "driver" is given in Figure 1, with English glosses added. The underlined material was not in Lexeed originally, we add it in this paper. *doraibā* "driver" has a familiarity of 6.55, and three senses. Lexeed has 28,000 words divided into 46,000 senses and defined with 75,000 definition sentences.

Useful hypernym relations can also be extracted from large corpora using relatively simple patterns (e.g., [Pantel *et al.*, 2004]). However, even a large newspaper corpus does not include all the familiar words of a language, let alone those words occurring in useful patterns [Amano and Kondo, 1999]. Therefore it makes sense to extract data from machine readable dictionaries (MRDs).

HEADWORD	ドライバー *doraiba-*		
POS	noun	<u>Lexical-type</u>	`noun-lex`
FAMILIARITY	6.5 [1–7]		
SENSE 1	DEFINITION	S_1	ねじ/まわし/。 screw turn (screwdriver)
		$S_1{}'$	ねじ/を/差し入れ/たり/、/抜き取っ/た/する/<u>道具</u>/。 A <u>tool</u> for inserting and removing screws .
	<u>HYPERNYM</u>	道具$_1$ *equipment* "tool"	
	<u>SEM. CLASS</u>	⟨**942:tool**⟩ (⊂ **893:equipment**)	
	<u>WORDNET</u>	*screwdriver*$_1$ (⊂ *tool*$_1$)	
SENSE 2	DEFINITION	S_1	自動車/を/運転/する/<u>人</u>/。 <u>Someone</u> who drives a car
	<u>HYPERNYM</u>	人$_1$ *hito* "person"	
	<u>SEM. CLASS</u>	⟨**292:driver**⟩ (⊂ **4:person**)	
	<u>WORDNET</u>	*driver*$_1$ (⊂ *person*$_1$)	
SENSE 3	DEFINITION	S_1	ゴルフ/で/、/遠/距離/用/の/<u>クラブ</u>/。 In golf, a long-distance <u>club</u>.
		S_2	一番/ウッド/。/ A number one wood .
	<u>HYPERNYM</u>	クラブ$_2$ *kurabu* "club"	
	<u>WORDNET SENSE</u>	*driver*$_5$ (⊂ *club*$_5$)	
	<u>DOMAIN</u>	ゴルフ$_1$ *gorufu* "golf"	

Figure 1: Entry for the Word *doraiba-* "driver" from Lexeed (with English glosses)

```
hook(h1)
    proposition_m_rel(h1,h3:)
        qeq(h3:,h17)
    _jidousha_n(h4,x5:)
    udef_rel(h6,x5:)
        RSTR(h6,h7:)
        BODY(h6,h8:)
        qeq(h7:,h4)
    _unten_s_2(h9,e11:present:)
        ARG1(h9,x10:)
        ARG2(h9,x5:)
    _hito_n(h12,x10:)
        ING(h12:,h10001:)
    udef_rel(h13,x10:)
        RSTR(h13,h14:)
        BODY(h13,h15:)
        qeq(h14:,h12)
    proposition_m_rel(h10001,h16:)
        qeq(h16:,h9)
    unknown_rel(h17,e2:present:)
        ARG2(h17,x10:)
```

RMRS from JACY (deep)

```
hook(h9)

    _jidousha_n(h1,x2)
    o_rel(h3,u4)

    _unten_s(h5,e6)
    suru_rel(h7,e8)

    _hito_n(h9,x10)
```

RMRS from ChaSen (shallow)

jidōsha wo unten suru hito ''a person who drives a car (lit: car-ACC drive do person)''

Real predicates are shown in **bold** font.

Figure 2: Deep and Shallow RMRS results for *doraibā*$_2$

2.2 Parsing Resources

We used the robust minimal recursion semantics (RMRS) designed in the Deep Thought project Callmeier *et al.* [2004], with tools from the Deep Linguistic Processing with HPSG Initiative (DELPH-IN: `http://www.delph-in.net/`).

Robust Minimal Recursion Semantics

Robust Minimal Recursion Semantics is a form of flat semantics which is designed to allow deep and shallow processing to use a compatible semantic representation, while being rich enough to support generalized quantifiers [Frank, 2004]. The full representation is basically the same as minimal recursion semantics [Copestake *et al.*, 2003]: a bag of labeled elementary predicates and their arguments, a list of scoping constraints, and a handle that provides a hook into the representation. The main difference is that handles must be unique, and there is an explicit distinction between grammatical and real predicates.

The representation can be underspecified in three ways: relationships can be omitted (such as message types, quantifiers and so on); predicate-argument relations can be omitted; and predicate names can be simplified. Predicate names are defined in such a way as to be as compatible as possible among different analysis engines (e.g. lemma-pos-sense, where sense is optional and the part of speech (pos) is drawn from a small set of general types (**n**oun, **v**erb, **s**ahen (verbal noun, ...)). The predicate **`unten_s`** is less specific than **`unten_s_2`** and thus subsumes it. In order to simplify the combination of different analyses, the results are indexed to the position in the original input sentence.

Examples of deep and shallow results for the same sentence 自動車を運転する人*jidōsha wo unten suru hito* "a person who drives a car (lit: car-ACC drive do person)" are given in Figure 2 (omitting the indexing). Real predicates are prefixed by an underbar (_). The deep parse gives information about the scope, message types and argument structure, while the shallow parse gives little more than a list of real and grammatical predicates with a hook.

Deep Parser (JACY and PET)

The Japanese grammar JACY [Siegel and Bender, 2002] was run with the PET System for the high-efficiency processing of typed feature structures [Callmeier, 2002].

Shallow Parser (based on ChaSen)

The part-of-speech tagger, ChaSen [Matsumoto *et al.*, 2000] was used for shallow processing of Japanese. Predicate names were produced by transliterating the pronunciation field and mapping the part-of-speech codes to the RMRS super types. The part-of-speech codes were also used to judge whether predicates were real or grammatical. Since Japanese is a head-final language, the hook value was set to be the handle of the rightmost real predicate.

3 Ontology Construction

As outlined in Section 1, our approach to ontology construction is to process a definition sentence with shallow and deep parsers and extract ontological relations from the most informative RMRS output. Here, we describe the algorithm used to extract ontological relations from an RMRS structure:

1. let P_i be the number of real predicates in the defining sentence
   - IF $P_i = 1$ (there is a unique real predicate) return: ⟨**`synonym`**: headword, predicate⟩
2. Initialize a stack of semantic relations to be processed with the semantic relation from the defining sentence's HOOK (the highest scoping handle)
3. Pop a semantic relation from the stack and check it against special predicates that require additional processing
   - When a relation indicating coordination or conjunction is found, locate all of its arguments and push them onto the stack for processing
   - IF a special predicate is found, extract its relations and add them to the stack
   - ELSE IF the current semantic relation is a real predicate, add it to list of extracted semantic heads

   Repeat until stack is empty
4. Return the ontological relations in the list of extracted semantic heads in the form: ⟨**`relation`**: headword, semantic_head⟩

Step 1. checks for a synonym relation, shown by a defining sentence containing a genus term with no differentia. Such a sentence will have a semantic representation with only a single real predicate.

In Step 2., for more complicated defining sentences, we try and find the genus term, looking first at the predicate with the widest scope. This is given by the RMRS's HOOK. The default ontological relation for the genus term is a hypernym.

Step 3. processes each semantic relation in the stack by searching for special predicates that require additional processing in order to retrieve the semantic head. Special predicates include explicit relation names (such as *ryaku* "abbreviation") and some grammatical predicates. This step identifies and processes special predicates, adding any results to the stack of unprocessed semantic relations. If a relation is not identified as being a special predicate, and it is a non-grammatical predicate, then it is accepted as a semantic head, and it is added to the list of extracted relations. Step 3 is repeated until the stack is empty.

Special predicates also give information about type of ontological relation that has been identified. They can confirm an implicit hypernym such as with *isshu* "one type" in Japanese or identify an entirely different relation, as in the case of the relation *part*, which identifies a meronym relationship in English or *meisho* "honorific name" identifying a name relation in Japanese. Specials predicates can also extract non-ontological relations such as domain.

Step 4. returns the list of all non-grammatical predicates once all semantic heads have been processed for special relations and no new results are produced.

This processing is following in the long tradition of parsing such special relationships (also called "empty heads", "function nouns" or "relators") [Guthrie *et al.*, 1990; Wilkes *et al.*, 1996]. Our main innovation is to extract them from the semantic representations produced by a deep and shallow

Special predicate	Ontological relation
Japanese	
isshu_n_1	hypernym
hitotsu_n_2	hypernym
soushou_n_1	hyponym
ryakushou_s_1	abbreviation
ryaku_s_1	abbreviation
keishou_n_1	name:honorific
zokushou_n_1	name:slang
meishou_n_1	name
bubun_n_1	meronym
ichibu_n_1	meronym

Table 1: Special predicates and their associated ontological relations

parsing, rather than using regular expressions or specially designed parsers.

4 Results and Evaluation

We summarize the relationships acquired in Table 2. The first two lines show thesaurus type relations: hypernyms and synonyms. The remaining four lines show other relations: abbreviations, names, meronyms and domains. Hypernyms and synonyms are by far the most common relations: fewer than 10% of entries are marked with an explicit relationship.

Results are shown for Lexeed using only the RMRS produced by ChaSen, using the results for JACY, and using the deepest possible result (JACY if it exists, backing off to ChaSen).

Results for ChaSen

Relation	Noun	Sahen	Verb	Other	Total
hypernym	42,235	8,176	9,237	3,346	62,994
synonym	7,278	776	2,005	933	10,992
Total	49,513	8,952	11,242	4,279	73,986

Results for JACY

Relation	Noun	Sahen	Verb	Other	Total
hypernym	31,374	6,748	6,619	2,029	46,770
synonym	7,831	801	2,220	1,048	11,900
abbreviation	154	7			161
domain	392	28			420
other	247				247
Total	39,998	7,584	8,839	3,077	59,498

Results for Deepest

Relation	Noun	Sahen	Verb	Other	Total
hypernym	45,014	9,647	10,305	3,299	68,265
synonym	81,51	827	2,257	1,254	12,489
abbreviation	154	7			161
domain	392	28			420
other	247				247
Total	53,958	10,509	12,562	4,553	81,582

Table 2: Results of Ontology Extraction (Lexeed)

As one would expect, the word based analysis using ChaSen finds more actual relationships, but does not provide enough information to find anything beyond implicit hypernyms and synonyms. The grammar based analyses have lower coverage, but allow us to extract some of the knowledge given explicitly in the lexicon.

Although the vast majority of relations extracted are hypernym and synonym relations, we extract several other kinds of knowledge, and are thus are closer to an ontology than a simple thesaurus.

We carried out two evaluations. The first was an automatic evaluation, comparing our extracted triples ⟨**`relation`**: word1, word2⟩ with existing resources. The second was a small scale hand evaluation of a sample of the relations.

4.1 Verification with Hand-crafted Ontologies

Because we are interested in comparing lexical semantics across languages, we compared the extracted ontology with resources in both the same and different languages.

We verified our results by comparing the hypernym links to the manually constructed Japanese ontology **GT**. It is a hierarchy of 2,710 semantic classes, defined for over 264,312 nouns [Ikehara *et al.*, 1997]. The semantic classes are mostly defined for nouns (and verbal nouns), although there is some information for verbs and adjectives. Senses are linked to **GT** semantic classes by the following heuristic: look up the semantic classes C for both the headword (w_i) and genus term(s) (w_g). If at least one of the index word's classes is subsumed by at least one of the genus' classes, then we consider the relationship confirmed (1).

$$\exists(c_h, c_g) : \{c_h \subset c_g; c_h \in C(w_h); c_g \in C(w_g)\} \quad (1)$$

In the event of an explicit hyponym relationship indicated between the headword and the genus, the test is reversed: we look for an instance of the genus' class being subsumed by the headword's class ($c_g \subset c_h$).

To test cross-linguistically, we looked up the headwords in a translation lexicon (**ALT-J/E** [Ikehara *et al.*, 1991] and EDICT [Breen, 2004]) and then did the confirmation on the set of translations $c_i \subset C(T(w_i))$. Although looking up the translation adds noise, the additional filter of the relationship triple effectively filters it out again.

For example, for ドライバー$_3$ *doraiba-*$_3$ "driver$_3$", **GT** does not find any relationship, as it does not have the golf club semantic class label for ドライバー*doraiba-*. However, looking up T(ドライバー) gives {*driver, screwdriver*} and the extracted hypernym is クラブ$_2$ *kurabu* "club". WordNet recognizes that *driver*$_5$ is a kind of *wood*$_8$ which is a kind of *club*$_5$ (using senses and relations from WordNet 2.0 [Fellbaum, 1998]). We thus simultaneously confirm the link is good; find an appropriate translation for this sense of ドライバー(and its hypernym); and link these to the appropriate WordNet synsems.

The results of the evaluation for Lexeed are shown in Table 3. Parts of speech are classified as "Noun", "Verb", "Sahen (Verbal Noun)" and "Other". Relations verified in either **GT** or WordNet are classed as verified. In these results, we only extract relations from the first definition sentence for each headword or when there is evidence of a **synonym** relation, as other definition sentences often provide clarification of the first sentence and may not contain useful ontological information. However, extracting relations from all definition sentences results in a net loss in confirmation of less than five percent while extracting over 5,000 additional relations. This suggests that even secondary definition sentences contain information that can be exploited in building ontologies.

Our results using JACY achieve a confirmation rate of **63.31%** for nouns only and **55.74%** overall, besting Tokunaga *et al.* [2001], who reported 61.4% for nouns only. Our method also allows us to extract multiple relations from a single sentence by processing coordinate clauses. This allowed us to extract an extra 3,300 relations. Using the deepest RMRS parse available, we confirmed **57.68%** of the noun relations and **50.49%** overall. This is comparable to our results reported in [Bond *et al.*, 2004] with one important difference: we are now extracting over 10,000 more confirmed ontological relations.

GT and WordNet both lack complete cover - over half the relations were confirmed with only one resource. This shows that the machine readable dictionary is a useful source of these relations. Cross lingual checking was surprisingly effective, resources in one language can be used to bootstrap those in another, as seen in the Euro WordNet project.

4.2 Human Evaluation

One problem with using existing ontological resources to verify new relations is that only relations which are subsumed by the ontology being used for comparison can be verified. This poses a considerable problem for researchers who wish to extract new relations: be it from domains where such resources are unavailable, or in cases where existing resources are limited in scope, such as for verbs. In this case, it makes more sense to evaluate a selection of the results retrieved by hand than to rely completely on existing ontologies for verification.

In this spirit, we conducted a hand-verification of a selection of our automatically acquired relations. 1,471 relations were selected using a stratified method over the entirety of our results (every 35th relationship, ordered by link-type and then headword). In this evaluation we only consider synonyms and any relationships extracted from the first sentence: the second and subsequent definition sentences tend to contain other non-hypernomic information. The results were then evaluated by native speakers of Japanese were given the definition word, the semantic head we retrieved, and the posited relation type and asked to evaluate if the relation was accurate. They had access to the original lexicon.

The human judges found the relations presented to them to be accurate 88.99% of the time. In the 162 relations that were judged unacceptable, it was also determined that a relation did exist in 95 cases, but it was incorrect (i.e. a **`synonym`** in place of a **`hypernym`** and so on). These errors had three sources: the most common was a lack of identified explicit relationships; the next was lack of information from the shallow parse and the last was errors in the argument structure of the deep parse. Tokunaga *et al.* [2001] report slightly higher results for extracting noun relationships only (91.8%).

5 Discussion and Future Work

We were able to successfully combine deep and shallow processing to extract ontological information from lexical resources. We showed that, by using a well defined semantic representation, the extraction can be generalized so much that it can be used on very different dictionaries from different languages. This is an improvement on the common approach to using more and more detailed regular expressions (e.g. Tokunaga *et al.* [2001]). Although this provides a quick start, the results are not generally reusable. In comparison, the ChaSen-RMRS engine is immediately useful for a variety of tasks, such as question answering and information extraction [Callmeier *et al.*, 2004]

The other innovation of our approach is the cross-lingual evaluation. As a by-product of the evaluation we enhance the existing resources (such as **GT** or WordNet) by linking them, so that information can be shared between them. Further, we hope to use the cross-lingual links to fill in gaps in the monolingual resources. Finally, we can trivially extract links from the **GT** ontology to WordNet, thus combining two useful resources and allowing us to compare them in detail.

There are several areas to address in future work as we continue to pursue ontology acquisition. First, and foremost, in order to increase the quality of the ontological relations that are extracted, we need to improve the quality of the parsers to generate RMRS. With our HPSG parsers, this can be addressed by adjusting the parse ranking model to take into account the special features of dictionary definition sentences. In addition we are currently increasing the coverage by adding in treatments of more grammatical phenomena.

Another area of great interest is the acquisition of other ontological relations. For example, if we extend our special predicates to include the relation produced by various forms of negation, we may be able to extract antonym relations.

Finally, we would also like to use the links created during evaluation which link our ontologies to hand-crafted ontologies, to furtherlink together senses of words across languages. This kind of cross-lingual ontology would be of great use in applications like machine translation.

6 Conclusion

We have demonstrated how deep and shallow processing techniques can be used together to enrich the acquisition of ontological information by constructing an ontology for Japanese. Our approach requires few rules and is thus easy to maintain and expand, and it can be easily extended to cover any language that has RMRS resources. In future research, we plan to extend our processing to retrieve more ontological relations and to investigate means of improving the accuracy of output of both deep and shallow processors.

References

[Amano and Kondo, 1999] Shigeaki Amano and Tadahisa Kondo. *Nihongo-no Goi-Tokusei (Lexical properties of Japanese)*. Sanseido, 1999.

[Bond *et al.*, 2004] Francis Bond, Eric Nichols, Sanae Fujita, and Takaaki Tanaka. Acquiring an ontology for a fundamental vocabulary. In *20th International Conference on Computational Linguistics: COLING-2004*, pages 1319–1325, Geneva, 2004.

Results for ChaSen

Relation	Noun (%)	Sahen (%)	Verb (%)	Other (%)	Total (%)
hypernym	13124 / 27779 (47.24)	2489 / 5856 (42.50)	2599 / 6903 (37.65)	397 / 2218 (17.90)	18609 / 42756 (43.52)
synonym	5684 / 7278 (78.10)	606 / 776 (78.09)	1285 / 2005 (64.09)	323 / 933 (34.62)	7898 / 10992 (71.85)
total	18808 / 35057 (53.65)	3095 / 6632 (46.67)	3884 / 8908 (43.60)	720 / 3151 (22.85)	26507 / 53748 (49.32)

Results for JACY

Relation	Noun (%)	Sahen (%)	Verb (%)	Other (%)	Total (%)
hypernym	12757 / 21634 (58.97)	2033 / 5130 (39.63)	1884 / 5254 (35.86)	376 / 1527 (24.62)	17050 / 33545 (50.83)
synonym	6099 / 7831 (77.88)	626 / 801 (78.15)	1351 / 2220 (60.86)	360 / 1048 (34.35)	8436 / 11900 (70.89)
abbreviation	61 / 149 (40.94)	3 / 7 (42.86)	–/– (–)	–/– (–)	64 / 156 (41.03)
domain	68 / 344 (19.77)	7 / 28 (25.00)	–/– (–)	–/– (–)	75 / 372 (20.16)
other	125 / 225 (55.56)	–/– (–)	–/– (–)	–/– (–)	125 / 225 (55.56)
total	19110 / 30183 (63.31)	2669 / 5966 (44.74)	3235 / 7474 (43.28)	736 / 2575 (28.58)	25750 / 46198 (55.74)

Results for Deepest

Relation	Noun (%)	Sahen (%)	Verb (%)	Other (%)	Total (%)
hypernym	15703 / 29731 (52.82)	2723 / 7141 (38.13)	2762 / 7927 (34.84)	492 / 2350 (20.94)	21680 / 47149 (45.98)
synonym	6307 / 8151 (77.38)	643 / 827 (77.75)	1371 / 2257 (60.74)	409 / 1254 (32.62)	8730 / 12489 (69.90)
abbreviation	61 / 149 (40.94)	3 / 7 (42.86)	–/– (–)	–/– (–)	64 / 156 (41.03)
domain	68 / 344 (19.77)	7 / 28 (25.00)	–/– (–)	–/– (–)	75 / 372 (20.16)
other	125 / 225 (55.56)	–/– (–)	–/– (–)	–/– (–)	125 / 225 (55.56)
total	22264 / 38600 (57.68)	3376 / 8003 (42.18)	4133 / 10184 (40.58)	901 / 3604 (25.00)	30674 / 60391 (50.79)

Table 3: Results confirmed for Lexeed (for 46,000 senses)

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

[Callmeier *et al.*, 2004] Ulrich Callmeier, Andreas Eisele, Ulrich Schäfer, and Melanie Siegel. The DeepThought core architecture framework. In *Proceedings of LREC-2004*, volume IV, Lisbon, 2004.

[Callmeier, 2002] Ulrich Callmeier. Preprocessing and encoding techniques in PET. In Stephan Oepen, Dan Flickinger, Jun-ichi Tsujii, and Hans Uszkoreit, editors, *Collabarative Language Engineering*, chapter 6, pages 127–143. CSLI Publications, Stanford, 2002.

[Copestake *et al.*, 2003] Ann Copestake, Daniel P. Flickinger, Carl Pollard, and Ivan A. Sag. Minimal Recursion Semantics. An introduction. 2003.

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

[Frank, 2004] Anette Frank. Constraint-based RMRS construction from shallow grammars. In *20th International Conference on Computational Linguistics: COLING-2004*, pages 1269–1272, Geneva, 2004.

[Guthrie *et al.*, 1990] Louise Guthrie, Brian M. Slator, Yorick Wilkes, and Rebecca Bruce. Is there content in empty heads? In *13th International Conference on Computational Linguistics: COLING-90*, pages 138–143, Helsinki, 1990.

[Ikehara *et al.*, 1991] Satoru Ikehara, Satoshi Shirai, Akio Yokoo, and Hiromi Nakaiwa. Toward an MT system without pre-editing – effects of new methods in **ALT-J/E**–. In *Third Machine Translation Summit: MT Summit III*, pages 101–106, Washington DC, 1991. (`http://xxx.lanl.gov/abs/cmp-lg/9510008`).

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

[Kasahara *et al.*, 2004] Kaname Kasahara, Hiroshi Sato, Francis Bond, Takaaki Tanaka, Sanae Fujita, Tomoko Kanasugi, and Shigeaki Amano. Construction of a Japanese semantic lexicon: Lexeed. SIG NLC-159, IPSJ, Tokyo, 2004. (in Japanese).

[Matsumoto *et al.*, 2000] Yuji Matsumoto, Kitauchi, Yamashita, Hirano, Matsuda, and Asahara. *Nihongo Keitaiso Kaiseki System: Chasen*, version 2.2.1 manual edition, 2000. url=http://chasen.aist-nara.ac.jp.

[Pantel and Lin, 2000] Patrick Pantel and Dekang Lin. An unsupervised approach to prepositional phrase attachment using contextually similar words. In *38th Annual Meeting of the Association for Computational Linguistics: ACL-2000*, pages 101–108, Hong Kong, 2000.

[Pantel *et al.*, 2004] Patrick Pantel, Deepak Ravichandran, and Eduard Hovy. Towards terascale knowledge acquisition. In *20th International Conference on Computational Linguistics: COLING-2004*, pages 771–777, Geneva, 2004.

[Siegel and Bender, 2002] Melanie Siegel and Emily M. Bender. Efficient deep processing of Japanese. In *Procedings of the 3rd Workshop on Asian Language Resources and International Standardization at the 19th International Conference on Computational Linguistics*, Taipei, 2002.

[Tokunaga *et al.*, 2001] Takenobu Tokunaga, Yasuhiro Syotu, Hozumi Tanaka, and Kiyoaki Shirai. Integration of heterogeneous language resources: A monolingual dictionary and a thesaurus. In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium, NLPRS2001*, pages 135–142, Tokyo, 2001.

[Tsurumaru *et al.*, 1991] Hiroaki Tsurumaru, Katsunori Takesita, Itami Katsuki, Toshihide Yanagawa, and Sho Yoshida. An approach to thesaurus construction from Japanese language dictionary. In *IPSJ SIGNotes Natural Language*, volume 83-16, pages 121–128, 1991. (in Japanese).

[Wilkes *et al.*, 1996] Yorick A. Wilkes, Brian M. Slator, and Louise M. Guthrie. *Electric Words*. MIT Press, 1996.