BaseNP Supersense Tagging for Japanese Texts

Hirotoshi Taira, Sen Yoshida, and Masaaki Nagata

NTT Communication Science Laboratories,
2-4, Hikaridai, Seika-cho, Keihanna Science City, Kyoto 619-0237, Japan
{taira, yoshida}@cslab.kecl.ntt.co.jp
nagata.masaaki@lab.ntt.co.jp

Abstract. This paper describes baseNP supersense tagging for Japanese texts. The task extracts base noun phrases (baseNPs) from raw texts in Japanese, and labels their baseNPs with supersenses. This task has a number of applications including predicate argument structure analysis and question answering. While the definition of baseNP in English is relatively clear, its definition in Japanese has not yet been clear. In this paper, we defined Japanese baseNP analogous to English and defined Japanese supersenses using a broad-coverage Japanese thesaurus, *Nihongo Goi Taikei* (comprehensive outline of Japanese vocabulary). We then adopted a sequential tagging algorithm for the task, namely the averaged perceptron with HMM, and achieve high performance compared to a baseline.

Keywords: Supersense, BaseNP, Named Entity, Predicate Argument Structure Analysis, Semantic Role Labeling.

1 Introduction

Named entity recognition (NER), has been useful for various natural language processing tasks such as searching for answer candidates in factoid question answering systems. However, if the answer is a common noun, NER cannot recognize the answer candidate. Ciaramita et al. proposed supersense tagging for noun phrases including common nouns and showed that the task has many applications (Ciaramita and Johnson, 2003) (Ciaramita and Altun, 2006).

Moreover, predicate argument structure analysis has attracted the attention of researchers recently because this information can increase the precision of text processing tasks, such as machine translation, information extraction (Hirschman et al., 1999), question answering (Shen and Lapata, 2007), and summarization (Melli et al., 2005). In the analysis, it is necessary to determine argument candidates, i.e. argument base noun phrases (baseNPs) before determining the semantic role of the candidates (Pradhan et al., 2004), and high performance noun phrase chunking is expected. Furthermore, supersenses annotated for NPs are helpful for predicate argument structure analysis (Taira et al., 2008), because we can use the case frame of verbs with semantic categories such as the NTT pattern pair dictionary (Fujita and Bond, 2008) and the large-scale case frame dictionary from the web (Kawahara and Kurohashi, 2006). Although baseNP chunking is a basic task in English and there are a lot of researches, the concept of baseNP in Japanese has been unclear. We propose a definition of Japanese baseNP in this paper.

We show the difference between NER and baseNP supersense tagging in Figure 1. Suppose that the sentence "彼は 5 日に記者会見を開いた。(He held a press interview on the 5th.)" is entered to the system. In this case, while NER only detects the 'TIME' phrase as '5 日 (the 5th)', the baseNP supersense tagger can recognize the common noun phrases, '彼 (he)' and '記者会見 (press interview)' as 'HUMAN' and 'HUMAN ACTIVITY', respectively. In the figure, 'O' stands

for ‘Other’, namely non-NP. Moreover, we adopted *Nihongo Goi Taikei* (comprehensive outline of Japanese vocabulary), whose coverage for nouns in Japanese is supposed to be the largest, as supersenses because the more the supersense of each word is defined in the lexicon, the better the performance of the baseNP supersense tagging.

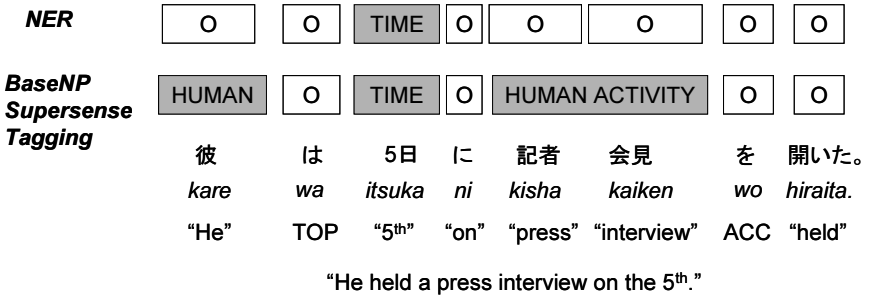

Figure 1: NER vs BaseNP Supersense Tagging

The rest of this paper is organized as follows. We describe a proposed definition of Japanese baseNP in Section 2. Next, we describe the *Nihongo Goi Taikei* and the supersenses defined on it in Section 3. We describe the algorithm for baseNP supersense tagging as a sequential labeling task in Section 4. In Section 5, we show our experiments and results. Our conclusions are provided in Section 6.

2 BaseNP in Japanese

The baseNP in English is defined as non-recursive noun phrase, i.e., a noun phrase not containing other noun phrase (Church, 1988) (Ramshaw and Marcus, 1995). We consider *Bunsetsu* phrases in Japanese excluding predicates (predicate bunsetsu phrases) as a possible candidate of the definition of the baseNP in Japanese. Bunsetsu phrase is a phonological unit of Japanese, containing one content word. However, Bunsetsu phrases often contain functional words and the meanings of more than two phrases sometimes change from that of the base phrase. So, we defined a definition of baseNP in Japanese as below.

1. Word sequence in phrases (*Bunsetsu* in Japanese) obtained by morphological analysis, excluding functional words that at the end of the last Bunsetsu.
2. However, if the supersense predicted by the individual words is different from the supersense of the entire noun phrase, take the shortest word sequence keeping the entire supersense.
3. As for relational clauses introduced by formal nouns (‘こと (thing)’, ‘の (that clause)’, etc.), take the formal noun (similar to the relative pronoun in English) as baseNP and label the formal noun with the supersense for the clause.

As for the first definition above, we could define shorter baseNP, namely a head word for a noun phrase. However, we cannot use the head word directly in many cases including answers in question answering and arguments in predicate argument analysis, because the head word in Japanese often does not have literal meaning. For example, the head word ‘者 (person)’ in the word ‘被害者 (victim),’ is usually used a suffix standing for a person and is not used the word itself.

The second definition includes proper nouns and idiomatic phrases. For example, a title of the movie, ‘ローマの休日 (Roman holiday)’ consists of two phrases in Japanese, namely ‘ローマの (of Rome)’ and ‘休日 (holiday),’ and the supersense of ‘休日 (holiday)’ itself is different from the supersense of the entire phrase. On the other hand, ‘楽しい休日 (delightful holiday)’ also consists of two phrases, namely ‘楽しい (delightful)’ and ‘休日 (holiday),’ and the supersense of ‘休日 (holiday),’ ‘TIME’ is the same as the supersense of the entire word, ‘楽しい休日 (delightful holiday),’ and we take ‘休日 (holiday)’ as the baseNP. If we want to use more informative expression,

‘楽しい休日 (delightful holiday)’ in some applications, in place of ‘休日 (holiday),’ we can also utilize dependency information such that the phrase ‘楽しい (delightful)’ depends on the phrase ‘休日 (holiday)’ and can use longer NPs.

The third definition avoids needlessly long baseNPs. For example, the sentence, ‘彼が手を振っていることに私は気がつかなかった (I did not notice that he was waving to me)’ contains a clause ‘彼が手を振っていること (that he was waving to me).’ The clause can be divided to ‘彼が手を振っている (he was waving to me)’ and a formal noun ‘こと (“that” clause marker).’ We label the formal noun ‘こと (that clause marker)’ with the supersense of ‘彼が手を振っていること (that he was waving for to me),’ namely ‘HUMAN ACTIVITY.’

The relation between the supersenses for a baseNP and each word in the baseNP is three-fold (Fig 2).

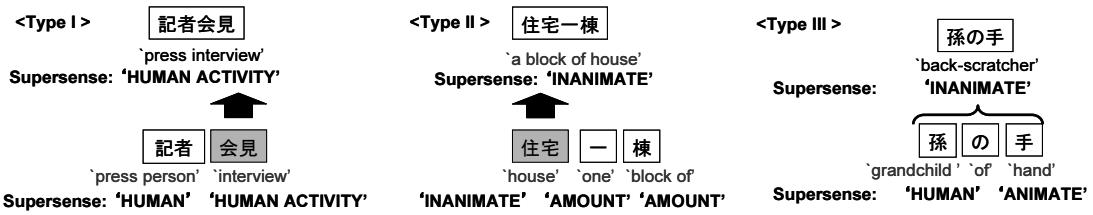

Figure 2: Three types of the relation between a supersense of a baseNP and words in the baseNP

1. The supersense for a baseNP is derived from the supersense of the last word in the baseNP.
2. The supersense for a baseNP is derived from the supersense of a word other than the last word in the baseNP.
3. The supersense for a baseNP is not derived from any words in the baseNP.

The existence of these different types makes the baseNP superesense tagging in Japanese difficult.

3 Nihongo Goi Taikei and Supersenses

WordNet (Fellbaum, 1998) is often used for supersenses in English (Ciaramita and Johnson, 2003). WordNet is a machine-readable dictionary and organized as a network of lexicalized concepts, sets of synonyms, called *synset.* Each noun synset can be assigned one out of 26 broad categories, called ‘supersenses.’ (Ciaramita and Johnson, 2003)

In Japanese, we used a well-known Japanese thesaurus, *Nihongo Goi Taikei* (Comprehensive outline of Japanese vocabulary), for supersenses. *Nihongo Goi Taikei* was originally developed for a Japanese-to-English machine translation system, ALT-J/E. It has three different semantic category hierarchies for common nouns, proper nouns, and verbs. Only the common noun category is widely used. The thesaurus consists of a hierarchy of 2,710 semantic classes, defined for over 264,312 nouns, with a maximum depth of twelve (Ikehara et al., 1997). The coverage for nouns are larger comparing with other Japanese thesaurus including *Bunrui Goi Hyo* (NIJL, 2004)(96,051 words) and Japanese WordNet (Isahara et al., 2008)(85,966 words (Ver.0.9)).

We used the semantic classes of the third level as supersenses because the level is similar to semantic roles. The top three levels of the Nihongo Goi Taikei common noun thesaurus are shown in Figure 3. For example, the Japanese word ライター (raitaa), which is derived from two different English words “writer” and “lighter”, but transliterated into the same Japanese string, has two different semantic categories, (353:author) and (915:household appliance). By following the *is-a* link, we can learn that the former sense refers to a person (4: person) while the latter sense refers to a physical object (706: inanimate object).

In the experiment, we also used the second level of Nihongo Goi Taikei for comparison. The thesaurus has 6 categories in the second level and 21 categories in the third level.

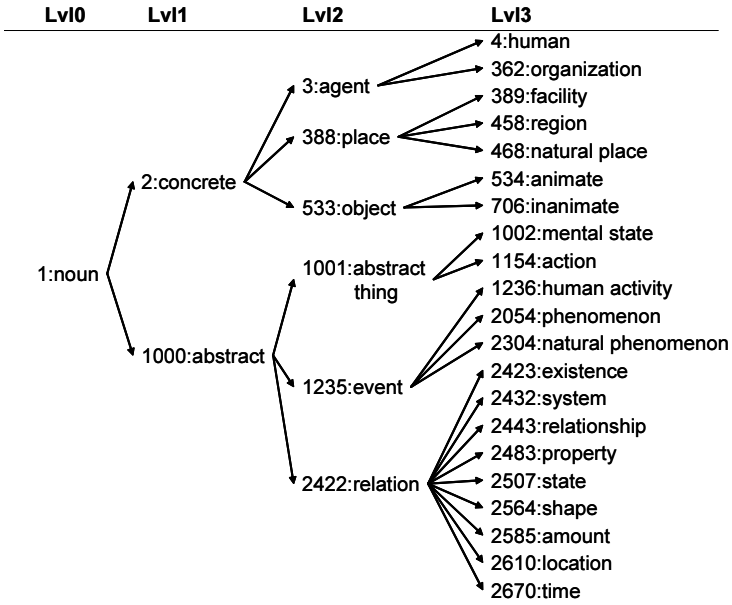

Figure 3: Top 3 levels of the Japanese thesaurus, 'Nihongo Goi Taikei'

4 BaseNP Supersense Tagging as Sequential Labeling

4.1 Averaged Perceptron with HMM

We used the averaged perceptron algorithm with HMM (Collins, 2002) for sequential tagging. Although a perceptron algorithm generally tends to overfit the training data, it avoids overfitting using a sort of voting method. The performance of the algorithm is reportedly comparable to that of the Conditional Random Fields (CRFs) (Sha and Pereira, 2003) and the calculation is generally faster and more memory efficient than that of CRFs.

The training algorithm is shown in Figure 4. Here, d is a document in a document set, s is a sentence in the document d, $x_{d,s,i}$ is the i-th word in the sentence s in the document d, T is the number of iterations, and w is a set of weights. $y^{gold}_{d,s,i}$ is the gold standard tag for $x_{d,s,i}$, and $y^{predicted}_{d,s,i}$ is the predicted tag by the system for $x_{d,s,i}$. $\Phi(x, y)$ is the feature set for (x, y). The final weight is calculated by averaging the weights after every iteration to avoid overfitting (Collins, 2002).

In the test phase, we calculate tag sequences with the maximum sum of weights w using a Viterbi algorithm and we have predicted tag sequences.

4.2 Features

We used the following binary features. We automatically segment a sentence into words and labeled parts of speech using Chasen (Matsumoto et al., 1997), which is a Japanese morphological analyzer. For training and test sets, we fixed the word regions and parts of speech manually to form training and test sets.

- Words ($wd_{-2}, wd_{-1}, wd_0, wd_{+1}, wd_{+2}$)
- POS ($pos_{-2}, pos_{-1}, pos_0, pos_{+1}, pos_{+2}$)
- POS First ($pos\_first_{-2}, pos\_first_{-1}, pos\_first_0, pos\_first_{+1}, pos\_first_{+2}$)
  The dictionary in Chasen uses a hierarchical part of speech system, and the first level refers to the major part of speech, such as noun, verb, etc.
- Supersense for Word (wd_sem_0)
  The first sense defined in Nihongo Goi Taikei is automatically annotated.

Input training samples $(x_{d,s,i}, y^{gold}_{d,s,i})$
Initialize $w_0 \leftarrow 0$
for (t = 0 to $T-1$) **do**
  for document d in Document Set **do**
    for sentence s in d **do**
      Compute $y^{predict}_{d,s,1}$ ... $y^{predict}_{d,s,max}$
        s.t. maximize $\sum w_t$
        for words in s, $x_{d,s,1}$ ... $x_{d,s,max}$ using Viterbi algorithm
      for $x_{d,s,i}$ in s **do**
        if $y^{gold}_{d,s,i} \neq y^{predict}_{d,s,i}$ **then**
          $w_{t+1} \leftarrow w_t + \Phi\ (\ x_{d,s,i}, y^{gold}_{d,s,i}\ ) - \Phi\ (\ x_{d,s,i}, y^{predict}_{d,s,i}\ )$
        end if
      end for
    end for
  end for
end for
$w = \frac{1}{T} \sum_t w_t$
Output w

Figure 4: Training algorithm

- Dependency (dep_0)
  The combination of functional words in the phrase containing the target word, and the head word in the phrase the target word depends on. The dependency analysis is obtained by Cabocha (Kudo and Matsumoto, 2003), which is a Japanese dependency analyzer, and we fixed the mistaken dependencies by hand.
- Next tag (y_{+1})
  We also used the predicted supersense of the next word.

4.3 Sequential Tag Format

We can understand the baseNP tagging task as a sequential labeling task (Ciaramita and Altun, 2006). There are some different formats for encoding chunks in the sequences (Sang and Veenstra, 1999) (Uchimoto et al., 2000). Kudo et al. indicated that the performance is the highest in a baseNP chunking task (not including supersense tagging) in English with SVM when they used the IOE2 format and the processing direction was backwards (Kudo and Matsumoto, 2002). This can probably be attributed to the fact that the head word at the chunk often exists in the end of the chunk. The situation is similar in Japanese, and we adopted the IOE2 format and backward processing. Figure 5 shows an example of the IOE2 tag format and a part of features in our task.

5 Experiments

5.1 Experimental Setting

We performed our experiments using Kyoto Corpus in 1995 (Mainichi, 1995), which is often used for evaluations of text processing in Japanese. We used articles published between January 1st and January 11th as training examples, and articles published between January 12th and 13th as test examples. We show the distribution of training and test data for the experiments in Table1.

5.2 Overall Results

First, we compared our system with a baseline method, which annotates a noun phrase with the supersense of the last word in the phrase.

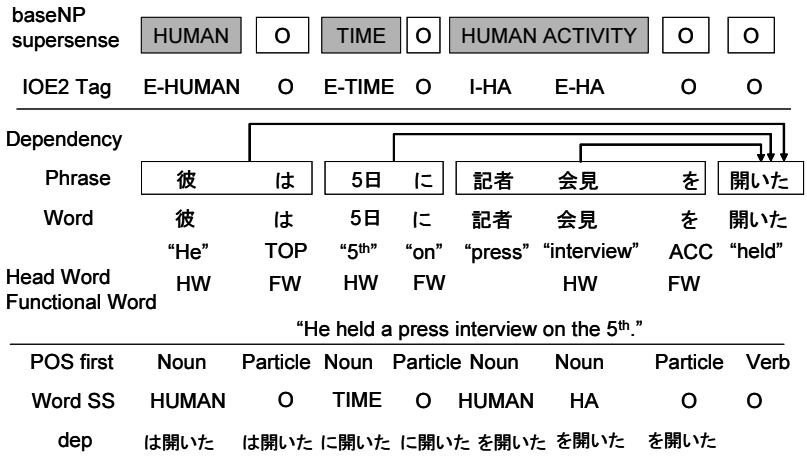

Figure 5: IOE2 tag format and features

Table 1: Distribution of training and test data.

	Training data	Test data
# articles	1,350	428
# sentences	11859	3,208
# words	324,792	91,145
# baseNPs	87,712	24,807

We show the results we obtained in Table 2. Here, ‘boundary’ indicates an evaluation of only the noun phrase boundary and ‘boundary+sem’ indicates an evaluation of the noun phrase boundary and supersense categories. And ‘AP with HMM’ stands for our system using the averaged perceptron algorithm with HMM. Our system is superior to the baseline system as regards both the second and third level supersenses.

Table 2: Comparison with baseline system (F-measure(%)).

	Level 2		Level 3	
mothod	baseline	AP with HMM	baseline	AP with HMM
boundary	89.13	**96.07**	89.13	**96.00**
boundary + sem	79.63	**89.33**	76.56	**86.79**

5.3 Effectiveness of dependency information

Next, we examined the effectiveness of the dependency information. We show the difference between the system performance with and without the ‘dep’ feature. We found that the feature is somewhat effective in both the second and third levels.

5.4 Effectiveness of multiple sense and immediate parent sense

We used only the first sense in our dictionary as the default word supersense. However, the dictionary also has some other senses if the word has multiple senses. So, we examined the feature set including all the senses in the dictionary. Moreover, we examined the effectiveness of the features using the parent nodes in the thesaurus. We show the results as for the third level superesenses in Tables 4. Here ‘m1’ and ‘m0’ indicate that the multiple senses were and were not used, respectively. ‘u1’ and ‘u2’ indicate the use of the parent nodes in the first and second levels, respectively.

Table 3: Effectiveness of dependency information (F-measure(%)).

	non dep (Level 2)	dep (Level 2)	non dep (Level 3)	dep (Level 3)
boundary	96.07	**96.21**	96.00	**96.08**
boundary + sem	89.33	**89.51**	86.79	**86.91**

The results indicate the limited effectiveness of both the multiple sense and parent node features as regards performance.

Table 4: Results for Test Data (Level 3)(F-measure(%)).

	m0	m0, u2	m0, u1	m1	m1, u2	m1, u1
boundary	96.08	**96.10**	96.07	96.09	96.04	95.81
boundary + sem	**86.91**	86.88	86.75	83.46	83.09	82.45

5.5 Effectiveness of tag format

Table 5 shows the results we obtained when we used IOE2 and IOB2 as the tag format. It can be seen that IOE2 tag format is greatly superior to the IOB2 tag format as we predicted.

Table 5: Effectiveness of tag format (F-measure(%)).

	IOB2 (Level 2)	IOE2 (Level 2)	IOB2 (Level 3)	IOE2 (Level 3)
boundary + sem	86.20	**89.51**	82.48	**86.91**

6 Conclusion

We described baseNP supersense tagging in Japanese. First, we defined a baseNP and supersenses in Japanese. Next, we adopt a sequential tagging algorithm for the task, namely an averaged perceptron with HMM, and a large semantic dictionary, and achieve a relatively high level of performance. This task has broad range of applications including predicate argument structure analysis and question answering.

References

Church, K. 1988. A stochastic parts program and noun phrase parser for unrestricted text. In *Proc. of the Second Conference on Applied Natural Language Processing (ANLP'88)*, pages 136–143.

Ciaramita, M. and Y. Altun. 2006. Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proc. of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, pages 594–602.

Ciaramita, M. and M. Johnson. 2003. Supersense tagging of unknown nouns in WordNet. In *Proc. of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP 2003)*, pages 168–175.

Collins, M. 2002. Discriminative training methods for hidden markov models. In *Proc. of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 1–8.

Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Fujita, S. and F. Bond. 2008. A method of creating new valency entries. *Machine Translation*, 21(1):1–28.

Hirschman, L., P. Robinson, L. Ferro, N. Chinchor, E. Brown, R. Grishman, and B. Sundheim. 1999. Hub-4 Event'99 general guidelines.

Ikehara, S., M. Miyazaki, S. Shirai, A. Yokoo, H. Nakaiwa, K. Ogura, Y. Ooyama, and Y. Hayashi. 1997. *Nihongo Goi Taikei, A Japanese Lexicon*. Iwanami Shoten, Tokyo.

Isahara, H., F. Bond, K. Uchimoto, M. Utiyama, and K. Kanzaki. 2008. Development of japanese wordnet. In *Proc. of the Sixth International Language Resources and Evaluation (LREC-2008)*, pages 2420–2423.

Kawahara, D. and S. Kurohashi. 2006. A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis. *Proc. of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL2006)*, pages 176–183.

Kudo, T. and Y. Matsumoto. 2002. Chunking with support vector machines (in Japanese). *Journal of Natural Language Processing*, 9(5):3–21.

Kudo, T. and Y. Matsumoto. 2003. Fast methods for kernel-based text analysis. In *Proc. of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, pages 24–31.

Mainichi. 1995. *CD Mainichi Shinbun 95*. Nichigai Associates Co.

Matsumoto, Y., A. Kitauchi, T. Yamashita, Y. Hirano, O. Imaichi, and T. Imamura, 1997. *Japanese Morphological Analysis System Chasen Manual*. NAIST Technical Report NAIST-IS-TR97007.

Melli, G., Y. Wang, Y. Liu, M. M. Kashani, Z. Shi, B. Gu, A. Sarkar, and F. Popowich. 2005. Description of SQUASH, the SFU question answering summary handler for the DUC-2005 summarization task. In *Proc. of DUC 2005*.

NIJL. 2004. *Bunrui Goi Hyo (in Japanese)*. Dainippon Tosho, Tokyo.

Pradhan, S., W. Ward, K. Hacioglu, J. Martin, and D. Jurafsky. 2004. Shallow semantic parsing using support vector machines. In *Proc. of the Human Language Technology Conference/North American Chapter of the Association of Computational Linguistics HLT/NAACL 2004*.

Ramshaw, L. A. and M. P. Marcus. 1995. Text chunking using transformation-based learning. In *Proc. of the Second Wworkshop on Very Large Corpora (WVLC'95)*, pages 82–94.

Sang, E. F. T. K. and J. Veenstra. 1999. Representing text chunks. In *Proc. of the Ninth Conference of the European Chapter of the ACL (EACL 99)*, pages 173–179.

Sha, F. and F. Pereira. 2003. Shallow parsing with conditional random fields. In *Proc. of HLT-NAACL 2003*, pages 213–220.

Shen, D. and M. Lapata. 2007. Using semantic roles to improve question answering. In *Proc. of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP/CoNLL)*, pages 12–21.

Taira, H., S. Fujita, and M. Nagata. 2008. A japanese predicate argument structure analysis using decision lists. In *Proc. of 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, pages 522–531.

Uchimoto, K., Q. Ma, M. Murata, H. Ozaku, M. Uchiyama, and H. Isahara. 2000. Named entity extraction based on a maximum entropy model and transformation rule (in Japanese). *Journal of Natural Language Processing*, 7(2):63–90.