

Reinforcing English Countability Prediction with One Countability per Discourse Property

Ryo Nagata

Hyogo University of Teacher Education
6731494, Japan
rnagata@hyogo-u.ac.jp

Atsuo Kawai

Mie University
5148507, Japan
kawai@ai.info.mie-u.ac.jp

Koichiro Morihiro

Hyogo University of Teacher Education
6731494, Japan
mori@hyogo-u.ac.jp

Naoki Isu

Mie University
5148507, Japan
isu@ai.info.mie-u.ac.jp

Abstract

Countability of English nouns is important in various natural language processing tasks. It especially plays an important role in machine translation since it determines the range of possible determiners. This paper proposes a method for reinforcing countability prediction by introducing a novel concept called *one countability per discourse*. It claims that when a noun appears more than once in a discourse, they will all share the same countability in the discourse. The basic idea of the proposed method is that mispredictions can be correctly overridden using efficiently the one countability per discourse property. Experiments show that the proposed method successfully reinforces countability prediction and outperforms other methods used for comparison.

1 Introduction

Countability of English nouns is important in various natural language processing tasks. It is particularly important in machine translation from a source language that does not have an article system similar to that of English, such as Chinese and Japanese, into English since it determines the range of possible determiners including articles. It also plays an important role in determining whether a noun can take singular and plural forms. Another useful application is to detect errors in article usage and singular/plural usage in the writing of second language learners. Given countability, these errors can be detected in many cases. For example, an error can be detected from “We have a furniture.” given that the noun *furniture* is un-

countable since uncountable nouns do not tolerate the indefinite article.

Because of the wide range of applications, researchers have done a lot of work related to countability. Baldwin and Bond (2003a; 2003b) have proposed a method for automatically learning countability from corpus data. Lapata and Keller (2005) and Peng and Araki (2005) have proposed web-based models for learning countability. Others including Bond and Vatikiotis-Bateson (2002) and O’Hara et al. (2003) use ontology to determine countability.

In the application to error detection, researchers have explored alternative approaches since sources of evidence for determining countability are limited compared to other applications. Articles and the singular/plural distinction, which are informative for countability, cannot be used in countability prediction aiming at detecting errors in article usage and singular/plural usage. Returning to the previous example, the countability of the noun *furniture* cannot be determined as uncountable by the indefinite article; first, its countability has to be predicted without the indefinite article, and only then whether or not it tolerates the indefinite article is examined using the predicted countability. Also, unlike in machine translation, the source language is not given in the writing of second language learners such as essays, which means that information available is limited.

To overcome these limitations, Nagata et al. (2005a) have proposed a method for predicting countability that relies solely on words (except articles and other determiners) surrounding the target noun. Nagata et al. (2005b) have shown that the method is effective to detecting errors in article usage and singular/plural usage in the writing of Japanese learners of English. They

also have shown that it is likely that performance of the error detection will improve as accuracy of the countability prediction increases since most of false positives are due to mispredictions.

In this paper, we propose a method for reinforcing countability prediction by introducing a novel concept called *one countability per discourse* that is an extension of *one sense per discourse* proposed by Gale et al. (1992). It claims that when a noun appears more than once in a discourse, they will all share the same countability in the discourse. The basic idea of the proposed method is that initially mispredicted countability can be corrected using efficiently the one countability per discourse property.

The next section introduces the one countability per discourse concept and shows that it can be a good source of evidence for predicting countability. Section 3 discusses how it can be efficiently exploited to predict countability. Section 4 describes the proposed method. Section 5 describes experiments conducted to evaluate the proposed method and discusses the results.

2 One Countability per Discourse

One countability per discourse is an extension of *one sense per discourse* proposed by Gale et al. (1992). One sense per discourse claims that when a polysemous word appears more than once in a discourse it is likely that they will all share the same sense. Yarowsky (1995) tested the claim on about 37,000 examples and found that when a polysemous word appeared more than once in a discourse, they took on the majority sense for the discourse 99.8% of the time on average.

Based on one sense per discourse, we hypothesize that when a noun appears more than once in a discourse, they will all share the same countability in the discourse, that is, one countability per discourse. The motivation for this hypothesis is that if one sense per discourse is satisfied, so is one countability per discourse because countability is often determined by word sense. For example, if the noun *paper* appears in a discourse and it has the sense of newspaper, which is countable, the rest of *papers* in the discourse also have the same sense according to one sense per discourse, and thus they are also countable.

We tested this hypothesis on a set of nouns¹

¹The conditions of this test are shown in Section 5. Note that although the source of the data is the same as in Section 5,

as Yarowsky (1995) did. We calculated how accurately the majority countability for each discourse predicted countability of the nouns in the discourse when they appeared more than once. If the one countability per discourse property is always satisfied, the majority countability for each discourse should predict countability with the accuracy of 100%. In other others, the obtained accuracy represents how often the one countability per discourse property is satisfied.

Table 1 shows the results. “MCD” in Table 1 stands for Majority Countability for Discourse and its corresponding column denotes accuracy where countability of individual nouns was predicted by the majority countability for the discourse in which they appeared. Also, “Baseline” denotes accuracy where it was predicted by the majority countability for the whole corpus used in this test.

Table 1: Accuracy obtained by Majority Countability for Discourse

Target noun	MCD	Baseline
advantage	0.772	0.618
aid	0.943	0.671
authority	0.864	0.771
building	0.850	0.811
cover	0.926	0.537
detail	0.829	0.763
discipline	0.877	0.652
duty	0.839	0.714
football	0.938	0.930
gold	0.929	0.929
hair	0.914	0.902
improvement	0.735	0.685
necessity	0.769	0.590
paper	0.807	0.647
reason	0.858	0.822
sausage	0.821	0.750
sleep	0.901	0.765
stomach	0.778	0.778
study	0.824	0.781
truth	0.783	0.724
use	0.877	0.871
work	0.861	0.777
worry	0.871	0.843
Average	0.851	0.754

Table 1 reveals that the one countability per discourses in which the target noun appears only once are excluded from this test unlike in Section 5.

course property is a good source of evidence for predicting countability compared to the baseline while it is not as strong as the one sense per discourse property is. It also reveals that the tendency of one countability per discourse varies from noun to noun. For instance, nouns such as *aid* and *cover* show a strong tendency while others such as *advantage* and *improvement* do not. On average, “MCD” achieves an improvement of approximately 10% in accuracy over the baseline.

Having observed the results, it is reasonable to exploit the one countability per discourse property for predicting countability. In order to do it, however, the following two questions should be addressed. First, how can the majority countability be obtained from a novel discourse? Since our intention is to predict values of countability of instances in a novel discourse, none of them are known. Second, even if the majority countability is known, how can it be efficiently exploited for predicting countability? Although we could simply predict countability of individual instances of a target noun in a discourse by the majority countability for the discourse, it is highly possible that this simple method will cause side effects considering the results in Table 1. These two questions are addressed in the next section.

3 Basic Idea

3.1 How Can the Majority Countability be Obtained from a Novel Discourse?

Although we do not know the true value of the majority countability for a novel discourse, we can at least estimate it because we have a method for predicting countability to be reinforced by the proposed method. That is, we can predict countability of the target noun in a novel discourse using the method. Simply counting the results would give the majority countability for it.

Here, we should note that countability of each instance is not the true value but a predicted one. Considering this fact, it is sensible to set a certain criterion in order to filter out spurious predictions. Fortunately, most methods based on machine learning algorithms give predictions with their confidences. We use the confidences as the criterion. Namely, we only take account of predictions whose confidences are greater than a certain threshold when we estimate the majority countability for a novel discourse.

3.2 How Can the Majority Countability be Efficiently Exploited?

In order to efficiently exploit the one countability per discourse property, we treat the majority countability for each discourse as a feature in addition to other features extracted from instances of the target noun. Doing so, we let a machine learning algorithm decide which features are relevant to the prediction. If the majority countability feature is relevant, the machine learning algorithm should give a high weight to it compared to others.

To see this, let us suppose that we have a set of discourses in which instances of the target noun are tagged with their countability (either *countable* or *uncountable*²) for the moment; we will describe how to obtain it in Subsection 4.1. For each discourse, we can know its majority countability by counting the numbers of *countables* and *uncountables*. We can also generate a model for predicting countability from the set of discourses using a machine learning algorithm. All we have to do is to extract a set of training data from the tagged instances and to apply a machine learning algorithm to it. This is where the majority countability feature comes in. The majority countability for each instance is added to its corresponding training data as a feature to create a new set of training data before applying a machine learning algorithm; then a machine learning algorithm is applied to the new set. The resulting model takes the majority countability feature into account as well as the other features when making predictions.

It is important to exercise some care in counting the majority countability for each discourse. Note that one countability per discourse is always satisfied in discourses where the target noun appears only once. This suggests that it is highly possible that the resulting model too strongly favors the majority countability feature. To avoid this, we could split the discourses into two sets, one for where the target noun appears only once and one for where it appears more than once, and train a model on each set. However, we do not take this strategy because we want to use as much data as possible for training. As a compromise, we approximate the majority countability for discourses where the target noun appears only once to the value *unknown*.

²This paper concentrates solely on countable and uncountable nouns, since they account for the vast majority of nouns (Lapata and Keller, 2005).

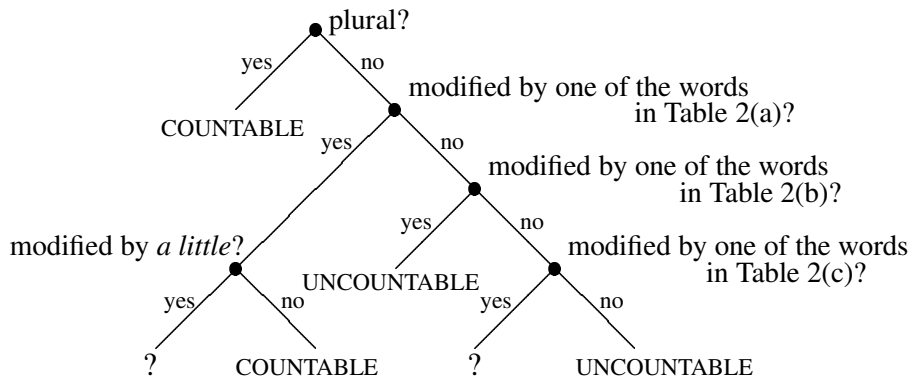


Figure 1: Framework of the tagging rules

(a)	(b)	(c)
<i>the indefinite article</i>	much	<i>the definite article</i>
another	less	<i>demonstrative adjectives</i>
one	enough	<i>possessive adjectives</i>
each	sufficient	<i>interrogative adjectives</i>
—	—	<i>quantifiers</i>
—	—	<i>'s genitives</i>

4 Proposed Method

4.1 Generating Training Data

As discussed in Subsection 3.2, training data are needed to exploit the one countability per discourse property. In other words, the proposed method requires a set of discourses in which instances of the target noun are tagged with their countability. Fortunately, Nagata et al. (2005b) have proposed a method for tagging nouns with their countability. This paper follows it to generate training data.

To generate training data, first, instances of the target noun used as a head noun are collected from a corpus with their surrounding words. This can be simply done by an existing chunker or parser.

Second, the collected instances are tagged with either *countable* or *uncountable* by tagging rules. For example, the underlined *paper*:

... read a paper in the morning ...

is tagged as

... read a paper/countable in the morning ...

because it is modified by the indefinite article.

Figure 1 and Table 2 represent the tagging rules based on Nagata et al. (2005b)’s method. Figure 1 shows the framework of the tagging rules. Each node in Figure 1 represents a question applied to the instance in question. For instance, the

root node reads “Is the instance in question plural?”. Each leaf represents a result of the classification. For instance, if the answer is “yes” at the root node, the instance in question is tagged with *countable*. Otherwise, the question at the lower node is applied and so on. The tagging rules do not classify instances in some cases. These unclassified instances are tagged with the symbol “?”. Unfortunately, they cannot readily be included in training data. For simplicity of implementation, they are excluded from training data (we will discuss the use of these excluded data in Section 6).

Note that the tagging rules cannot be used for countability prediction aiming at detecting errors in article usage and singular/plural usage. The reason is that they are useless in error detection where whether determiners and the singular/plural distinction are correct or not is unknown. Obviously, the tagging rules assume that the target text contains no error.

Third, features are extracted from each instance. As the features, the following three types of contextual cues are used: (i) words in the noun phrase that the instance heads, (ii) three words to the left of the noun phrase, and (iii) three words to its right. Here, the words in Table 2 are excluded. Also, function words (except prepositions) such as pronouns, cardinal and quasi-cardinal numer-

als, and the target noun are excluded. All words are reduced to their morphological stem and converted entirely to lower case when collected. In addition to the features, the majority countability is used as a feature. For each discourse, the numbers of *countables* and *uncountables* are counted to obtain its majority countability. In case of ties, it is set to *unknown*. Also, it is set to *unknown* when only one instance appears in the discourse as explained in Subsection 3.2.

To illustrate feature extraction, let us consider the following discourse (target noun: *paper*):

... writing a new *paper/countable* in his room ...
 ... read *papers/countable* with ...

The discourse would give a set of features:

-3=write, NP=new, +3=in, +3=room, MC=c
 -3=read, +3=with, MC=c

where “MC=c” denotes that the majority countability for the discourse is *countable*. In this example (and in the following examples), the features are represented in a somewhat simplified manner for the purpose of illustration. In practice, features are represented as a vector.

Finally, the features are stored in a file with their corresponding countability as training data. Each piece of training data would be as follows:

-3=read, +3=with, MC=c, LABEL=c

where “LABEL=c” denotes that the countability for the instance is *countable*.

4.2 Model Generation

The model used in the proposed method can be regarded as a function. It takes as its input a feature vector extracted from the instance in question and predicts countability (either *countable* or *uncountable*). Formally, $f : v \rightarrow c$ where f , v , and c denote the model, the feature vector, and $c = 0, 1$, respectively; here, 0 and 1 correspond to *countable* and *uncountable*, respectively.

Given the specification, almost any kind of machine learning algorithm can be used to generate the model used in the proposed method. In this paper, the Maximum Entropy (ME) algorithm is used which has been shown to be effective in a wide variety of natural language processing tasks.

Model generation is done by applying the ME algorithm to the training data. The resulting model takes account of the features including the majority countability feature and is used for reinforcing countability prediction.

4.3 Reinforcing Countability Prediction

Before explaining the reinforcement procedure, let us introduce the following discourse for illustration (target noun: *paper*):

... writing *paper* in room ... wrote *paper* in ...
 ... submitted *paper* to ...

Note that articles and the singular/plural distinction are deliberately removed from the discourse. This kind of situation can happen in machine translation from a source language that does not have articles and the singular/plural distinction³. The situation is similar in the writing of second language learners of English since they often omit articles and the singular/plural distinction or use improper ones. Here, suppose that the true values of the countability for all instances are *countable*.

A method to be reinforced by the proposed method would predict countability as follows:

... writing *paper/countable* (0.97) in room ...
 ... wrote *paper/countable* (0.98) in ...
 ... submitted *paper/uncountable* (0.57) to ...

where the numbers in brackets denote the confidences given by the method. The third instance is mistakenly predicted as *uncountable*⁴.

Now let us move on to the reinforcement procedure. It is divided into three steps. First, the majority countability for the discourse in question is estimated by counting the numbers of the predicted *countables* and *uncountables* whose confidences are greater than a certain threshold. In case of ties, the values of the majority countability is set to *unknown*. In the above example, the majority countability for the discourse is estimated to be *countable* when the threshold is set to 0.95 (two *countables*). Second, features explained in Subsection 4.1 are extracted from each instance. As for the majority countability feature, the estimated one is used. Returning to the above example, the three instances would give a set of features:

-3=write, +3=in, +3=room, MC=c,
 -3=write, +3=in, MC=c,
 -3=submit, +3=to, MC=c.

Finally, the model generated in Subsection 4.2 is applied to the features to predict countability. Because of the majority countability feature, it

³For instance, the Japanese language does not have an article system similar to that of English, neither does it mark the singular/plural distinction.

⁴The reason would be that the contextual cues did not appear in the training data used in the method.

is likely that previous mispredictions are overridden by correct ones. In the above example, the third one would be correctly overridden by *countable* because of the majority countability feature (MC=c) that is informative for the instance being *countable*.

5 Experiments

5.1 Experimental Conditions

In the experiments, we chose Nagata et al. (2005a)'s method as the one to be reinforced by the proposed method. In this method, the decision list (DL) learning algorithm (Yarowsky, 1995) is used. However, we used the ME algorithm because we found that the method with the ME algorithm instead of the DL learning algorithm performed better when trained on the same training data.

As the target noun, we selected 23 nouns that were also used in Nagata et al. (2005a)'s experiments. They are exemplified as nouns that are used as both countable and uncountable by Huddleston and Pullum (2002).

Training data were generated from the written part of the British National Corpus (Burnard, 1995). A text tagged with the text tags was used as a discourse unit. From the corpus, 314 texts, which amounted to about 10% of all texts, were randomly taken to obtain test data. The rest of texts were used to generate training data.

We evaluated performance of prediction by accuracy. We defined accuracy by the ratio of the number of correct predictions to that of instances of the target noun in the test data.

5.2 Experimental Procedures

First, we generated training data for each target noun from the texts using the tagging rules explained in Subsection 4.1. We used the OAK system⁵ to extract noun phrases and their heads. Of the extracted instances, we excluded those that had no contextual cues from the training data (and also the test data). We also generated another set of training data by removing the majority countability features from them. This set of training data was used for comparison.

Second, we obtained test data by applying the tagging rules described in Subsection 4.1 to each instance of the target noun in the 314 texts. Nagata et al. (2005b) showed that the tagging rules

⁵<http://www.cs.nyu.edu/~sekine/PROJECT/OAK/>

achieved an accuracy of 0.997 in the texts that contained no errors. Considering these results, we used the tagging rules to obtain test data. Instances tagged with “?” were excluded in the experiments.

Third, we applied the ME algorithm⁶ to the training data without the majority countability feature. Using the resulting model, countability of the target nouns in the test data was predicted. Then, the predictions were reinforced by the proposed method. The threshold to filter out spurious predictions was set to 0.95. For comparison, the predictions obtained by the ME model were simply replaced with the estimated majority countability for each discourse. In this method, the original predictions were used when the estimated majority countability was *unknown*. Also, Nagata et al. (2005a)'s method that was based on the DL learning algorithm was implemented for comparison.

Finally, we calculated accuracy of each method. In addition to the results, we evaluated the baseline on the same test data where all predictions were done by the majority countability for the whole corpus (training data).

5.3 Experimental Results and Discussion

Table 3 shows the accuracies⁷. “ME” and “Proposed” in Table 3 refer to accuracies of the ME model and the ME model reinforced by the proposed method, respectively. “ME+MCD” refers to accuracy obtained by replacing predictions of the ME model with the estimated majority countability for each discourse. Also, “DL” refers to accuracy of the DL-based method.

Table 3 shows that the three ME-based methods (“Proposed”, “ME”, and “ME+MCD”) perform better than “DL” and the baseline. Especially, “Proposed” outperforms the other methods in most of the target nouns.

Figure 2 summarizes the comparison between the three ME-based methods. Each plot in Figure 2 represents each target noun. The horizontal and vertical axes correspond to accuracy of “ME” and that of “Proposed” (or “ME+MCD”), respectively. The diagonal line corresponds to the line $y = x$. So if “Proposed” (or “ME+MCD”) achieved no improvement at all over “ME”, all the

⁶All ME models were generated using the `opennlp.maxent` package (<http://maxent.sourceforge.net/>).

⁷The baseline in Table 3 is different from that in Table 1 because discourses where the target noun appears only once are not taken into account in Table 1.

Table 3: Experimental results

Target noun	Freq.	Baseline	Proposed	ME	ME+MCD	DL
advantage	570	0.604	0.933	0.921	0.811	0.882
aid	385	0.665	0.909	0.873	0.896	0.722
authority	1162	0.760	0.857	0.851	0.840	0.804
building	1114	0.803	0.848	0.842	0.829	0.807
cover	210	0.567	0.790	0.757	0.800	0.714
detail	1157	0.760	0.906	0.904	0.821	0.869
discipline	204	0.593	0.804	0.745	0.750	0.696
duty	570	0.700	0.879	0.877	0.828	0.847
football	281	0.907	0.925	0.907	0.925	0.911
gold	140	0.929	0.929	0.929	0.921	0.929
hair	448	0.902	0.908	0.902	0.904	0.904
improvement	362	0.696	0.735	0.715	0.685	0.738
necessity	83	0.566	0.831	0.843	0.831	0.783
paper	1266	0.642	0.859	0.836	0.808	0.839
reason	1163	0.824	0.885	0.893	0.834	0.843
sausage	45	0.778	0.778	0.733	0.756	0.778
sleep	107	0.776	0.925	0.897	0.897	0.813
stomach	30	0.633	0.800	0.800	0.800	0.733
study	1162	0.779	0.832	0.819	0.782	0.808
truth	264	0.720	0.761	0.777	0.765	0.731
use	1390	0.869	0.879	0.863	0.871	0.873
work	3002	0.778	0.858	0.842	0.837	0.806
worry	119	0.798	0.874	0.840	0.849	0.849
Average	662	0.741	0.857	0.842	0.828	0.812

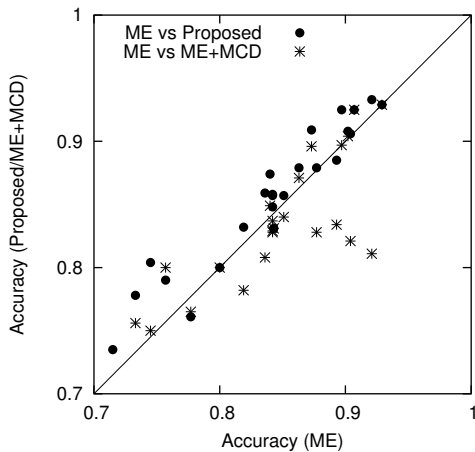


Figure 2: Comparison between “ME” and “Proposed/ME+MCD” in each target noun

plots would be on the line. Plots above the line mean improvement over “ME” and the distance from the line expresses the amount of improvement. Plots below the line mean the opposite.

Figure 2 clearly shows that most of the plots (•)

corresponding to the comparison between “ME” and “Proposed” are above the line. This means that the proposed method successfully reinforced “ME” in most of the target nouns. Indeed, the average accuracy of “Proposed” is significantly superior to that of “ME” at the 99% confidence level (paired *t-test*). This improvement is close to that of one sense per discourse (Yarowsky, 1995) (improvement ranging from 1.3% to 1.7%), which seems to be a sensible upper bound of the proposed method. By contrast, about half of the plots (*) corresponding to the comparison between “ME” and “ME+MCD” are below the line.

From these results, it follows that the one countability per discourse property is a good source of evidence for predicting countability, but it is crucial to devise a way of exploiting the property as we did in this paper. Namely, simply replacing original predictions with the majority countability for the discourse causes side effects, which has been already suggested in Table 1. This is

also exemplified as follows. Suppose that several instances of the target noun *advantage* appear in a discourse and that its majority countably is *countable*. Further suppose that an idiomatic phrase “take *advantage* of” of which countability is *uncountable* happens to appear in it. On one hand, simply replacing all the predictions with its majority countability (*countable*) would lead to a misprediction for the idiomatic phrase even if the original prediction is correct. On the other hand, the proposed method would correctly predict the countability because the contextual cues strongly indicate that it is *uncountable*.

6 Conclusions

This paper has proposed a method for reinforcing English countability prediction by introducing one countability per discourse. The experiments have shown that the proposed method successfully overrode original mispredictions using efficiently the one countability per discourse property. They also have shown that it outperformed other methods used for comparison. From these results, we conclude that the proposed method is effective in reinforcing English countability prediction.

In addition, the proposed method has two advantages. The first is its applicability. It can reinforce almost any earlier method. Even to hand-coded rules, it can be applied as long as they give predictions with their confidences. This further gives an additional advantage. Recall that the instances tagged with “?” by the tagging rules are discarded when training data are generated as described in Subsection 4.1. These instances can be retagged with their countability by using the proposed method and some kind of bootstrapping (Yarowsky, 1995). This means increase in training data, which might eventually result in further improvement. The second is that the proposed method is unsupervised. It requires no human intervention to reinforce countability prediction.

For future work, we will investigate what models are most appropriate for exploiting the one countability per discourse property. We will also explore a method for including instances tagged with “?” in training data by using the proposed method and bootstrapping.

Acknowledgments

The authors would like to thank Satoshi Sekine who has developed the OAK System. The authors

also would like to thank three anonymous reviewers for their useful comments on this paper.

References

- T. Baldwin and F. Bond. 2003a. Learning the countability of English nouns from corpus data. In *Proc. of 41st Annual Meeting of ACL*, pages 463–470.
- T. Baldwin and F. Bond. 2003b. A plethora of methods for learning English countability. In *Proc. of 2003 Conference on Empirical Methods in Natural Language Processing*, pages 73–80.
- F. Bond and C. Vatikiotis-Bateson. 2002. Using an ontology to determine English countability. In *Proc. of 19th International Conference on Computational Linguistics*, pages 99–105.
- L. Burnard. 1995. *Users Reference Guide for the British National Corpus, version 1.0*. Oxford University Computing Services, Oxford.
- W.A. Gale, K.W. Church, and D. Yarowsky. 1992. One sense per discourse. In *Proc. of 4th DARPA Speech and Natural Language Workshop*, pages 233–237.
- R. Huddleston and G.K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge.
- M. Lapata and F. Keller. 2005. Web-based models for natural language processing. *ACM Transactions on Speech and Language Processing*, 2(1):1–31.
- R. Nagata, F. Masui, A. Kawai, and N. Isu. 2005a. An unsupervised method for distinguishing mass and count noun in context. In *Proc. of 6th International Workshop on Computational Semantics*, pages 213–224.
- R. Nagata, T. Wakana, F. Masui, A. Kawai, and N. Isu. 2005b. Detecting article errors based on the mass count distinction. In *Proc. of 2nd International Joint Conference on Natural Language Processing*, pages 815–826.
- T. O’Hara, N. Salay, M. Witbrock, D. Schneider, B. Aldag, S. Bertolo, K. Panton, F. Lehmann, J. Curtis, M. Smith, D. Baxter, and P. Wagner. 2003. Inducing criteria for mass noun lexical mappings using the Cyc KB, and its extension to WordNet. In *Proc. of 5th International Workshop on Computational Semantics*, pages 425–441.
- J. Peng and K. Araki. 2005. Detecting the countability of English compound nouns using web-based models. In *Companion Volume to Proc. of 2nd International Joint Conference on Natural Language Processing*, pages 105–109.
- D. Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proc. of 33rd Annual Meeting of ACL*, pages 189–196.