Crosslingual Countability Classification with EuroWordNet

Leonoor van der Beek and Timothy Baldwin

Groningen University   Stanford University

Abstract

We examine the hypothesis that noun countability is consistent for a given word semantics by way of a series of experiments involving **EuroWordNet** and the English and Dutch languages. The basic method involves determining a default set of countabilities for each **EuroWordNet** synset based on countability-mapped words in that synset, and testing the match between these countabilities and those of held-out words. As **EuroWordNet** provides crosslingual synset correspondences between Dutch and English, we are able to evaluate the method both monolingually for Dutch and English, and crosslingually between the two languages. We found that Dutch and English countabilities align as well cross-lingually as they do monolingually.

1 Introduction

Ontologies such as **WordNet**[1] (Fellbaum 1998) and **EuroWordNet**[2] (Vossen 1998) comprise a hierarchical network of concept nodes, populated with words. The nodes in **WordNet**-style networks are conventionally termed **synsets**, as they contain sets of synonymous words representing a common underlying concept. Synsets offer a means of semantic generalization, both over the component words within a given synset and between synsets (and by extension their component words) via hierarchical relations such as hyponymy (subordination) and hypernymy (superordination). These forms of generalization have been successfully applied in a variety of tasks including text categorization (e.g. de Buenaga Rodríguez, Gómez Hidalgo and Díaz Agudo (2000)), PP attachment (e.g. (Stetina and Nagao 1997)), subcategorization frame acquisition (e.g. Preiss, Korhonen and Briscoe (2002)), selectional preference learning (e.g. Clark and Weir (2002)) and information retrieval (e.g. Mandala, Tokunaga and Tanaka (2000))

This paper examines the use of synsets in the automatic acquisition of lexical properties of individual words. The underlying assumption is that some lexical properties are not (completely) arbitrary, but to a large extent determined by semantics, and moreover that **WordNet** synsets are at an appropriate level of semantic granularity to capture such properties. Under this assumption, the determination of lexical properties can be made at the synset level and applied to the individual members through simple propagation. Determination of synset-level properties is possible by inheriting the lexical properties of annotated members of a given synset.

The particular lexical property we focus on in this research is noun countability. In lexical terms, the countability class of a noun governs determiner cooccurrence, the ability to pluralize, and enumeration effects. There are conflicting

[1] `http://www.cogsci.princeton.edu/~wn/`
[2] `http://www.illc.uva.nl/EuroWordNet/`

claims as to the semantic grounding of countability (Wierzbicka 1988, Jackendoff 1991, Gillon 1996), but in terms of lexical ontologies, previous research has shown there to be a high correlation between the synset membership of English nouns and their countability classification (Bond and Vatikiotis-Bateson 2002, O'Hara, Salay, Witbrock, Schneider, Aldag, Bertolo, Panton, Lehmann, Smith, Baxter, Curtis and Wagner 2003). We take this line of research a step further in exploring the possibilities for mono- and crosslingual ontology-based countability classification in both English and Dutch, using **EuroWordNet** as our common resource. That is, we attempt to determine the countability of each synset in **EuroWordNet** from Dutch and/or English training data, and then evaluate the accuracy of the synset-level countability predictions over held-out data in the two languages.

English and Dutch are closely-related languages and the basic nature of noun countability aligns well in the two languages. Both languages distinguish between the three countability classes of countable, uncountable and plural only,[3] and although mismatches exist—e.g. *hersenen* (plural only) vs. *brain* (countable), *onweer* (uncountable) vs. *thunderstorm* (countable)—many Dutch words are in the same countability class as their English equivalents (e.g. *fiets* / *bike*, *eten* / *food*, *goederen* / *goods*). Through direct comparison of monolingual and crosslingual classification, this research empirically quantifies the level of countability consistency between the two languages, relative to in-language consistency.

Knowledge of countability is important for both analysis and generation. In analysis it helps to reduce lexical and syntactic ambiguity. In generation, countability information determines whether a noun can be pluralized and what determiners it can combine with. Finally, countability information is required to distinguish syntactically-marked constructions such as determinerless PPs, involving countable nouns, from their unmarked counterparts, involving uncountable nouns (Baldwin, Beavers, van der Beek, Bond, Flickinger and Sag 2003).

In the following, we define countability and outline the lexical resources used in this research (Section 2), and describe previous research (Section 3). We then detail the classification procedures (Section 4) and evaluate each method (Section 5).

2 Preliminaries

2.1 Countability Classes

We consider both Dutch and English to have the three countability classes of countable (also known as "count"), uncountable (also known as "mass") and plural only.[4] **Countable** nouns can be modified by denumerators (prototypically numbers), and generally have a morphologically-marked plural form: *een fiets* / *one bike*, *twee fietsen* / *two bikes*. This class contains nouns which are easily individuated (i.e. there is a clear concept of a "base unit" of the concept). **Uncountable**

[3] A fourth class of bipartite nouns (e.g. *scissors*, *trousers*) is generally recognized for English, but has no Dutch correlate.
[4] Haeseryn et al. (1997) use a slightly different ontology: "uncountable" is used as an umbrella term for *pluralia tantum* (our plural only) and *singularia tantum* (our uncountable).

nouns cannot be modified by denumerators, do not have a plural form, but can be modified by unspecific quantifiers such as *veel* / *much*: **een eten* / *one food*, *een beetje eten* / *some food*, **twee etens* / *two foods*. This class includes many abstract, material-denoting, collective and deverbalized nouns. **Plural only** nouns have only a plural form, and cannot be denumerated: *goederen* / *goods*. Since the plural only class is considered to be a closed class in Dutch, the classification experiments below focus exclusively on the countable and uncountable classes, ignoring plural only nouns.

It is important to realize that different senses/usages of a given word can occur with different countabilities, cf. *Ik wil een konijn* / *I want a rabbit* (countable) vs. *I zou graag nog wat konijn willen* / *I would like some more rabbit, please* (uncountable). It is not necessarily the case, however, that because a given word occurs with distinct countabilities it has multiple senses. Consider, e.g., *I ordered a beer* (countable) vs. *I ordered beer* (uncountable), which we claim correspond to a single sense of *beer*.

Accounts of countability range from a purely semantically motivated feature (Jackendoff 1991) to a completely arbitrary lexical feature in many computational grammars, e.g. the Alpino grammar (Bouma, van Noord and Malouf 2001). The former runs into problems when faced with different realizations of one concept in different languages, such as the Dutch *onweer* vs. English *thunderstorm*, mentioned earlier. The latter fails to account for the semantic underpinnings and crosslingual commonalities of countability. Moreover, it implies that type-level countability distinctions are categorical, which is in fact not the case (Allan 1980): prototypical countable nouns can be used in uncountable contexts, forcing a ‘substance’ interpretation (the **universal grinder**, e.g. *over de hele straat lag hert* / *there was deer all over the road*) and uncountable nouns can be denumerated in certain contexts, resulting in a ‘type’ interpretation (the **universal packager**, e.g. *deze winkel verkoopt drie verschillende wijnen* / *this shop sells three different wines*). This being said, nouns are generally considered to have a predominant use or basic classification as countable and/or uncountable. Copestake (1992) accounts for both the arbitrary aspects and conversion. The semantic types **countable** and **uncountable** are used to capture the default classification and lexical rules are provided to account for conversion from one type to the other.

Following Bond and Vatikiotis-Bateson (2002) and O’Hara et al. (2003), we assume that the countability of a noun is to a large extent predictable from its semantic class, or in this case its synset. By experimenting with different classification methods, we hope to filter out the ‘noise’ introduced by nouns that have an unpredictable and language-specific countability and capture the generalization that countability is stable for a given word semantics.

2.2 Lexical Resources

Information about English noun countability was obtained from two lexical sources: **COMLEX** 3.0 (Grishman, Macloed and Myers 1998) and the common noun part of **ALT-J/E**’s Japanese-to-English semantic transfer dictionary

Language	Dataset	Size	EWN mapped	Mean EWN polysemy	Agreement (%)
English (EN)	$\text{Dictionary}_{\text{EN}}$	5,853	5,826	2.1	85.6
	$\text{Learned}_{\text{EN}}$	11,357	6,974	1.5	82.0
	$\text{Dic+Learn}_{\text{EN}}$	17,210	12,800	1.8	83.8
	$\text{Annotated}_{\text{EN}}$	98	70	1.5	—
Dutch (NL)	$\text{Dictionary}_{\text{NL}}$	14,400	10,407	1.9	81.1
	$\text{Learned}_{\text{NL}}$	5,819	2,213	1.8	85.7
	$\text{Dic+Learn}_{\text{NL}}$	19,661	12,088	1.9	82.4
	$\text{Annotated}_{\text{NL}}$	196	159	2.0	—
Dutch & English	$\text{Comb}_{\text{EN/NL}}$	36,871	24,888	1.8	83.1

Table 1: Countability datasets

(Bond 2001). These two resources were combined by taking the intersection of positive and negative exemplars for each countability class. The total number of training instances is around 6,000 words; we refer to this dataset as **Dictionary**$_{\text{EN}}$ for the remainder of this paper. To evaluate the quality of our data, we hand-annotated 100 unseen nouns according to actual usage in the British National Corpus (BNC: Burnard (2000)) and measured the agreement[5] with Dictionary$_{\text{EN}}$ to be 85.6%.

In addition to the dictionary data, we used a second data set consisting of some 34,000 nouns that were automatically classified on the basis of corpus data (Baldwin and Bond (2003a, 2003b)). In Section 3, we describe the procedure that was used for the corpus based classification. From the classified nouns, we extracted the (countable and uncountable) common nouns, which numbered about 11,000 in total; we refer to this dataset as **Learned**$_{\text{EN}}$. We once again hand-annotated 100 nouns from this set according to actual usage in the BNC, to make up dataset **Annotated**$_{\text{EN}}$. The agreement between Annotated$_{\text{EN}}$ and Learned$_{\text{EN}}$ is 82.0%.

In evaluation, we use the combination of Dictionary$_{\text{EN}}$ and Learned$_{\text{EN}}$, making up 17,000 English nouns at 83.8% agreement with the gold standard (**Dic+Learn**$_{\text{EN}}$).

We have access to an analogous set of countability datasets in Dutch, based on dictionary, learned and manually-annotated data. The dictionary data (**Dictionary**$_{\text{NL}}$) was extracted from the **Alpino** lexicon (Bouma et al. 2001) and represents a manually-modified and extended version of the countability data found in **CELEX**. The total number of Dutch nouns is around 14,500, of which some 10,400 were listed in **EuroWordNet**.

The learned Dutch countability data is based on the method described in Baldwin and van der Beek (2003), as applied to around 6,000 common nouns not found in the **Alpino** lexicon. The learning method combines corpus-based and word-to-word classification methods (see Section 3) based on both English and Dutch

[5] I.e. the proportion of word-level countability class assignments over which the two sets agreed.

training data. We refer to the resultant dataset as **Learned**$_{\rm NL}$ hereafter.

As with English, we will exclusively use the combination of these datasets in evaluation, which we will refer to as **Dic+Learn**$_{\rm NL}$.

In order to test the quality of the Dutch dictionary-derived and learned data, we manually annotated 196 unseen Dutch nouns, basing judgments on actual usage in the Twente Nieuws Corpus;[6] we refer to this dataset as **Annotated**$_{\rm NL}$. The agreement in countability judgments between Dictionary$_{\rm NL}$ and Annotated$_{\rm NL}$ is 81.1%, somewhat lower than for the English dictionary data; the agreement for Learned$_{\rm NL}$ was a more respectable 85.7%, almost identical to that for Dictionary$_{\rm EN}$.

Finally, we combined the two English datasets with the Dutch Alpino data to form a single multilingual dataset of about 37,000 countability-classified nouns at overall agreement of 83.1%, which we label as **Comb**$_{\rm EN/NL}$. For an overview of the datasets, see Table 1.

We used **EuroWordNet** to determine the synset membership of a given noun, and also to map Dutch and English synsets onto one another. Three components were used: the Dutch database of nouns, the English database of nouns and the Inter-Lingual Index (ILI). The Dutch component contains about 35,000 nouns, grouped into synsets. The English component is a reformatted version of **WordNet** 1.5, and contains nearly 88,000 nouns. The ILI interconnects the monolingual ontologies by way of hyponym, hypernym, synonym and near-synonym relations. Each record in the ILI is in turn connected to the **WordNet** 1.5 ontology by way of one or more "offsets", each representing a **WordNet** synset. Multiple offsets are used to collapse portions of the **WordNet** 1.5 structure which correspond to systematic polysemy or overly fine-grained sense distinctions, and also to add sense distinctions which are made in two or more of the languages targeted by **EuroWordNet** but not in the original **WordNet** 1.5 ontology.

In Table 1, we present the number of nouns in each dataset which is mapped onto the EWN ontology, and also the mean polysemy of each **EuroWordNet**-mapped noun (i.e. the average number of senses per noun). We observe that there is very little difference in the total number of **EuroWordNet**-mapped nouns in the Dic+Learn$_{\rm EN}$ and Dic+Learn$_{\rm NL}$ datasets, and that mean polysemy is almost identical in the two languages. When combined with our estimated values for agreement with the annotated data, the two datasets are thus remarkably similar in size and overall quality, allowing us to make the claim that any biases observed in evaluation is intrinsic in the languages rather than being peculiar to the datasets.

3 Past Research

Past research on countability classification falls into three basic categories: corpus-based, concept-based and word-to-word.

Corpus-based countability classification is based on the premise that the countability of a word type is reflected in its corpus token occurrences, in the form of co-occurrence patterns (e.g. with determiners, verbs or prepositions). Baldwin and Bond (2003a, 2003b) applied this approach to the task of English countability

[6] `http://wwwhome.cs.utwente.nl/~druid/TwNC/TwNC-main.html`

classification in two forms: (a) distribution-based classification, which is based on the relative frequency of different features over token occurrences of a given word; and (b) agreement-based classification, which uses the output of multiple pre-processors to measure the degree of token agreement over features found to uniquely correlate with a given countability class. Distribution-based classification thus looks for feature distribution "signatures" characteristic of different countabilities, whereas agreement-based classification looks for convincing evidence of occurrence of one or more features which are strong indicators of a given countability. In evaluation over the four countability classes of countable, uncountable, plural only and bipartite using BNC data, they found distribution-based classification to be the superior method, achieving 94.6% agreement with dictionary data (or 89.2% agreement for only the countable and uncountable classes). It is this method of distribution-based countability classification that we use in generating the learned English dataset described in Section 2.2.

Schwartz (2002) also performed corpus-based countability classification, constructing an automatic countability tagger (ACT) to learn token-level noun countabilities from the BNC. The method has a coverage of around 50%, and agrees with **COMLEX** for 68% of the nouns marked countable and with the **ALT-J/E** lexicon for 88%.

Baldwin and van der Beek (2003) used corpus-based countability classification in a crosslingual context, to learn Dutch countability based on Dutch and English corpus data, and English countability annotations. The crosslingual mapping takes the form of feature alignment between English and Dutch, either at the feature cluster level (e.g. all determiner features correlating with countable nouns are considered as a single whole) or at the individual feature level (e.g. the English *a* co-occurrence feature is mapped onto the Dutch *een* co-occurrence feature). By using the feature alignment schema to transform the feature vectors for English and Dutch noun types, Baldwin and van der Beek used the English feature vectors as training data in classifying the Dutch data.

Concept-based countability classification—as employed in this research—is based on the assumption that members of a given concept class or synset have the same countability. It has been applied to English by Bond and Vatikiotis-Bateson (2002) using the **ALT-J/E** ontology, and O'Hara et al. (2003) using the **Cyc** ontology and English **WordNet**. Bond and Vatikiotis-Bateson cite an accuracy of 78% over a 5-way classification of countability preference, whereas O'Hara et al. achieve an accuracy of 89.5% over the two-way distinction of countable/uncountable using **Cyc**. We are unaware of any research which has attempted concept-based countability classification in a crosslingual context.

Word-to-word countability classification uses direct lexical alignment to determine the countability of novel words from corresponding countability-annotated words. Baldwin and van der Beek (2003) applied this strategy in a crosslingual context using English-to-Dutch word-to-word translation and transliteration data as the source of alignment. They found the method to be remarkably accurate, with transliteration data achieving an accuracy of 98.3%, but to have limited coverage.

4 Classifier Design

We experimented with classifiers that vary along two dimensions: the classification method and the **EuroWordNet** link types between training and test words. The classification methods we used are union-based classification, majority-based classification and combined classification. The **EuroWordNet** link types we experimented with are (near-)synonyms, hypernyms and hyponyms. In our first set of experiments, we test the different classification methods over (near-)synonym training words only. In a second set of experiments, we then include countability information from hypernyms and hyponyms.

While we have acknowledged that different senses of a word can occur with different countabilities, we have no immediate way of determining which **EuroWordNet** senses of a given word correspond to which countability.[7] We are thus forced to assign the countability class(es) of each noun to all its senses in **EuroWordNet**.

4.1 Classification method

In this section, we detail each of the classification methods proposed in this research. We illustrate their differences by way of the Dutch noun *wederpartij* "antagonist/adversary" (countable) and the English-to-Dutch crosslingual classification task, using the Dictionary$_{\mathrm{EN}}$ dataset. In **EuroWordNet**, *wederpartij* maps onto **WordNet** offsets 6071277 (glossed as "a hostile person who tries to do damage to you") and 5922580 (glossed as "someone who offers opposition"). English nouns mapped onto **WordNet** offset 6071277 are *opponent* (countable), *opposition* (uncountable) and *enemy* (countable), with the indicated countabilities in the dictionary dataset; English nouns mapped onto **WordNet** offset 5922580 are *adversary*, *antagonist* and *opponent*, of which the dictionary dataset lists only *opponent* as countable. In our discussion of each classification method, we discuss how this countability information is used in classifying *wederpartij*.

Union-based classification

For each target noun, the union-based classifier determines the countability class(es) of all training words occurring in the synset(s) of the target noun. The noun is then assigned the union of all attested countability classes.

Under this method, *wederpartij* is classified as being both countable (by virtue of its similarity to *enemy* and *opponent*) and uncountable (by virtue of its similarity to *opposition*).

[7] In fact, countabilities in the **ALT-J/E** lexicon are tailored to the different senses of each word, but given our partial use of its countability data and the lack of an established mapping between the **ALT-J/E** ontology and **EuroWordNet** synsets, we are unable to make use of this information.

Majority-based classification

Majority-based classification is based on simple voting between the countability classes of the training words in the relevant synset(s). The target noun is assigned the (unique) most frequently attested countability class, and in the case of a tie, defaults to countable.

Under majority-based classification, *wederpartij* receives three votes for countable and one vote for uncountable, and is thus classified as being countable.

Combined classification

The combined classifier maps nouns to countability classes in two steps. First, it uses majority-based classification to determine a unique classification within each synset. It then takes the union of the individual synset-based classifications. This reflects the intuition that the different countability classifications for a word are often related to the different senses of that lexical item. Also, the combined classifier is designed to filter out low-frequency countabilities in each synset a given word occurs in, hence reducing the effect of language-specific, unpredictable countability mappings of training words.

In the case of *wederpartij*, both **WordNet** synsets receive a countable classification, leading to the final classification of countable.

4.2 EuroWordNet link type

Synonym-based classification

In synonym-based classification, we completely ignore the hierarchical structure of the ILI and use it as a simple sense inventory, expanding out each ILI record into its corresponding **WordNet** offset(s) (= synsets). In the crosslingual case, therefore, we end up with synsets comprising nouns in both Dutch and English.

The countability of each target noun is determined on the basis of the countability classes of those words occurring in the same **WordNet** synset(s), following one of the three classification methods described above.

Hypernym-based classification

We also experimented with hypernym-based countability classification. The underlying (simplifying) assumption is that traversing a hypernymy link (i.e. traversing up the **WordNet** hierarchy) does not change the countability, and so the hypernyms can be used as additional training data in countability classification.

Classification takes place according to two steps: (1) we first look for synonyms of the target word in the training data, and if found, perform synonym-based classification; (2) failing this, we use the ILI to identify hypernym synsets of the different senses of the word, and base the class determination on training data in hypernym synsets.

Hyponym-based classification

Hyponym-based classification is similar to hyponym-based classification. The only difference is that we traverse down rather than up the **WordNet** hierarchy via hyponym links in the second classification step, and base the countability classification on the countabilities of hyponym words.

Bidirectional classification

Bidirectional classification combines hypernym- and hyponym-based classification, and in the second step of classification looks both up and down the **EuroWordNet** hierarchy, basing classification on the combination of hypernyms and hyponyms.

We expect that the inclusion of hypernyms and hyponyms in the set of training words will lead to higher coverage (i.e. we will be able to find at least one countability for more words). On the other hand, we expect mismatches in countability to arise more frequently, e.g. *tafel* / *table* (countable) vs. its hypernym *meubilair* / *furniture* (uncountable).

5 Results and Discussion

In this section we present the results for the various classification methods using each **EuroWordNet** link type, over different combinations of training and test datasets. We start with a basic comparison of the results for the different classification methods based on synonymy (Section 5.1), and classify using the different **EuroWordNet** link types (Section 5.2). We then present a breakdown of the results over countable and uncountable nouns (Section 5.3), and finally contrast mono- and crosslingual classification (Section 5.4).

Classifier performance is rated according to precision (P), recall (R) and F-score (F). All calculations are based on test words which are contained in **EuroWordNet** and which have at least one countability-mapped training noun in one of the synsets accessed by the classification method in question.

Throughout evaluation, we use the combined dictionary and learned countability data for English and Dutch (i.e. Dic+Learn$_{\text{EN}}$ and Dic+Learn$_{\text{NL}}$) to classify nouns in both languages. That is, we make no distinction between crosslingual and monolingual countability classification.

5.1 Performance of each Classification Method

We first evaluate the three classification methods—union-based (Union), majority-based (Major) and combined (Comb)—according to synonym **EuroWordNet** links. The results are presented in Table 2, evaluated according to F-score. In the case that the training and test datasets are the same, evaluation is according to 10-fold cross-validation.

The results for majority-based classification and combined classification are on the whole markedly better than for union-based classification. The reason for the

TEST DATA	TRAINING DATA: Dic+Learn$_{EN}$ Union	Major	Comb	Dic+Learn$_{NL}$ Union	Major	Comb
Annotated$_{EN}$	.327	.423	**.469**	.510	.503	**.600**
Dic+Learn$_{EN}$	.401	.527	**.534**	.526	.610	**.616**
Annotated$_{NL}$	.646	.661	**.730**	.626	.598	**.658**
Dic+Learn$_{NL}$	.674	.713	**.719**	.674	.701	**.705**

Table 2: F-score using different classifi cation methods

TEST DATA	TRAINING DATA: Dic+Learn$_{EN}$ Syn	Hyper	Hypo	Both	Dic+Learn$_{NL}$ Syn	Hyper	Hypo	Both
Annotated$_{EN}$	.469	.802	.636	**.829**	.600	.811	.671	**.834**
Dic+Learn$_{EN}$	.531	.734	.614	**.747**	.616	.734	.641	**.737**
Annotated$_{NL}$	.730	**.802**	.762	.801	.658	**.783**	.690	.779
Dic+Learn$_{NL}$	.719	**.781**	.730	.770	.706	**.798**	.716	.783

Table 3: F-score using different **EuroWordNet** link types

generally poor performance of union-based classification as compared to the other methods appears to be its susceptibility to low-frequency noise and our inability to determine sense–countability correspondence. For example, *bloem* "flour/flower" (countable/uncountable) occurs in two synsets: "pulverized grain" and "bloom or blossom of a plant". The former of these corresponds to the uncountable sense and the latter to the countable sense, but as we lack such information, both synsets are wrongly predicted to be both countable and uncountable. In addition, some noise is introduced by arbitrary, unpredictable and language-particular countability mappings. As these are not correlated with the semantics of a word, they do not provide any information about the countability class of other synset members. This noise is filtered out by majority voting in majority-based and combined classification, but not in union-based classification.

Majority-based and combined classification are relatively close in performance, but combined classification returns a superior F-score over all combinations of training and test data. As a result, we use the combined method for the remainder of evaluation.

5.2 Performance of each EuroWordNet Link Type

We next evaluate the performance of combined classification making fuller use of the **EuroWordNet** hierarchy, through synonym (Syn), hypernym (Hyper), hy-

Test data	Dic+Learn$_{EN}$ Syn P	R	Hyper P	R	Hypo P	R	Dic+Learn$_{NL}$ Syn P	R	Hyper P	R	Hypo P	R
Annotated$_{EN}$	**1.00**	.312	.962	**.797**	.971	.516	**1.00**	.547	.952	**.938**	.975	.609
Dic+Learn$_{EN}$	**.887**	.427	.813	**.811**	.841	.541	**.840**	.577	.775	**.864**	.807	.627
Annotated$_{NL}$	.924	.713	.881	**.926**	**.931**	.794	.903	.684	.857	**.971**	**.909**	.735
Dic+Learn$_{NL}$	**.901**	.703	.840	**.930**	.884	.783	**.868**	.709	.822	**.969**	.856	.769

Table 4: Precision (P) and recall (R) for COUNTABLE nouns over different link types

Test data	Dic+Learn$_{EN}$ Syn P	R	Hyper P	R	Hypo P	R	Dic+Learn$_{NL}$ Syn P	R	Hyper P	R	Hypo P	R
Annotated$_{EN}$	**1.00**	.294	.857	**.529**	.789	.441	**1.00**	.206	.684	**.382**	.857	.353
Dic+Learn$_{EN}$	**.703**	.274	.505	**.550**	.583	.430	**.727**	.258	.513	**.430**	.579	.361
Annotated$_{NL}$	**.846**	.373	.588	**.508**	.574	.525	**.571**	.203	.529	**.305**	.486	.305
Dic+Learn$_{NL}$	**.478**	.453	.376	**.604**	.363	.561	**.520**	.283	.410	**.418**	.369	.385

Table 5: Precision (P) and recall (R) for UNCOUNTABLE nouns over different link types

ponym (Hypo) and bidirectional (Both) links. The results are given in Table 3, based on the same combination of datasets as in Table 2. The single best F-score for each training–test dataset combination is presented in **boldface**.

The use of both hypernym and hyponym links leads to an appreciable gain in F-score (relative to simple synonym data), and the combination of the two in the form of bidirectional classification tends to lead to a slight increment over the best of the individual methods (although it is interesting that the bidirectional method is marginally inferior to simple hypernyms for the Dutch data). Using information from hierarchically-linked synsets, or in other words, inheriting countabilities from words which are immediately superordinate or subordinate to the target noun appears to be a successful strategy in countability classification.

In general, hypernym-based classification was superior to hyponym-based classification. We explore the reason for this in the following section.

5.3 Performance over Countable and Uncountable Nouns

In an attempt to cast light on the relative performance of synonym-, hypernym- and hyponym-based classification, and their impact on each of the two countability classes, we broke down the results into precision (P) and recall (R), as presented in Tables 4 and 5. As above, the single best precision and recall value for each training–test dataset combination is presented in **boldface**.

Synonym-based classification tends to get the best precision, and hypernym-based classification by far the best recall. It is the big jump in recall afforded by hypernym-based classification that gives it the edge in terms of F-score, as seen above. The strong performance for synonym-based classification in terms of precision suggests that countability is more consistent within synsets than between adjacent synsets (i.e. synsets linked by hyponymy or hypernym links); it also justifies our cascaded approach to classification using **EuroWordNet** links (i.e. using hyponym and hypernym data only in the case that no synonym data is available). The reason for the disparity in recall between hyponym- and hypernym-based classification is that all synsets other than the root nodes have at least one hypernym, whereas only around 40% of the populated synsets in each dataset have hyponym synset(s) (i.e. are non-leaf nodes). The potential for hierarchical links to produce a countability judgment is thus considerably greater for hypernym-based classification.

The drop in precision between hyponym-based classification and hypernym-based classification does not represent a difference in the consistency of countability between the two link types. Indeed, evaluation of the two classification strategies independent of synonym-based classification suggested that hypernym-based classification offers marginally higher precision than hyponym-based classification. The difference in precision between the two strategies is thus due to the higher utility of hypernym-based classification impinging more noticeably on the base precision due to synonym-based classification.

Comparing countable and uncountable nouns, there is relatively little differential in precision (other than for Dutch uncountable nouns where both precision and recall drop considerably), but the recall for countable nouns is much higher. Looking at the actual occurrences of uncountable nouns in **EuroWordNet**, they appear to cluster less regularly along the vertical axis of the **EuroWordNet** hierarchy than countable nouns, for both English and Dutch, and certainly there are relatively few majority-uncountable classes. One possibility to gain extra leverage out of the **EuroWordNet** geometry which, based on observation, could be more effective in classifying uncountable nouns, would be to look at generalizing horizontally across the hierarchy via sister relations (a la Bond and Vatikiotis-Bateson (2002)). We leave this as an item for future research.

5.4 Mono- vs. Crosslingual Classification

To date, we have made nothing of the language combinations used in obtaining the results. In this section, we compare the relative performance of the crosslingual and monolingual classification tasks over the $\text{Annotated}_{\text{EN}}$ and $\text{Annotated}_{\text{NL}}$ datasets. The results are presented in Table 6, based on combined bidirectional classification.

One striking result observable in Table 6 is that crosslingual classification is superior to monolingual classification for both annotated datasets, i.e. we achieve the

TEST DATA	TRAINING DATA: Dic+Learn$_{EN}$			Dic+Learn$_{NL}$			Comb$_{EN/NL}$		
	P	R	F	P	R	F	P	R	F
Annotated$_{EN}$	**.904**	.765	.829	.876	.796	.834	.902	**.847**	**.874**
Annotated$_{NL}$	.752	.856	.801	.764	.795	.779	**.772**	**.867**	**.816**

Table 6: Performance over the annotated datasets (P = precision, R = recall, F = F-score)

best F-score for Dutch using English training data, and the best F-score for English using Dutch training data. That is, despite there being attested countability mismatches for conceptually-equivalent Dutch and English nouns (see Section 2.1), these appear to be no more pronounced than in-language countability inconsistencies for the two languages. This is a remarkable result, and underlines the linguistic similarity of the two languages. Note that while the differences in F-score in each case are relatively slim, when we generate a learning curve for the different combinations of training and test data, the margin is preserved remarkably consistently throughout.

The combination of the two monolingual training datasets in the form of Comb$_{EN/NL}$ performs significantly better over both test datasets, with particularly noticeable gains in recall. This appears to be a simple consequence of Comb$_{EN/NL}$ containing twice the number of training exemplars as the component datasets, and any advantage in classification performance is removed when we halve Comb$_{EN/NL}$ (to generate a dataset of equivalent size to Dic+Learn$_{EN}$ and Dic+Learn$_{NL}$).

6 Conclusion

We have presented several methods for applying **EuroWordNet** in automatic countability classification, relying on the semantic grounding of countability. The proposed methods varied on two dimensions: (1) the method used to formulate a countability judgment from the training data, and (2) what links we make use of within the **EuroWordNet** ontology in pooling together training data. We showed that it is possible to learn noun countability from conceptually-linked crosslingual data, using datasets from both Dutch and English. In doing so, we demonstrated empirically that Dutch and English countabilities align as well crosslingually as they do monolingually.

It is an interesting and yet unanswered question how this method would perform when applied to languages that are less closely related or differ with respect to the countability distinctions manifest in the languages. As the method is based only on conceptual similarity and draws its countability annotations from external sources, it can easily be applied to any language pair (assuming a common ontology and countability information in each language), even if there are divergences in the nature of countability in the two languages.

Acknowledgments

The authors would like to thank Francis Bond (NTT CS Labs.), the anonymous reviewers and also the audience of a presentation of an earlier version of this paper at the University of Groningen for valuable comments. Part of this research was carried out within the framework of the PIONIER Project *Algorithms for Linguistic Processing*, which is funded by NWO (Dutch Organization for Scientific Research) and the University of Groningen. The second author was supported by the National Science Foundation under Grant No. BCS-0094638 and also the Research Collaboration between NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation and CSLI, Stanford University.

References

Allan, K.(1980), Nouns and countability, *Language* **56**(3), 541–67.

Baayen, R. H., Piepenbrock, R. and van Rijn, H.(1993), *The CELEX Lexical Database (CD-ROM)*, Linguistic Data Consortium, University of Pennsylvania, Philadelphia, USA.

Baldwin, T. and Bond, F.(2003a), Learning the countability of English nouns from corpus data, *Proc. of the 41st Annual Meeting of the ACL*, Sapporo, Japan, pp. 463–70.

Baldwin, T. and Bond, F.(2003b), A plethora of methods for learning English countability, *Proc. of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP 2003)*, Sapporo, Japan, pp. 73–80.

Baldwin, T. and van der Beek, L.(2003), The ins and outs of Dutch noun countability classification, *Proceedings of the 2003 Australasian Language Technology Workshop (ALTW2003)*, Melbourne, Australia, pp. 33–40.

Baldwin, T., Beavers, J., van der Beek, L., Bond, F., Flickinger, D. and Sag, I. A.(2003), In search of a systematic treatment of determinerless PPs, *Proc. of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, Toulouse, France.

Bond, F.(2001), *Determiners and Number in English, contrasted with Japanese, as exemplified in Machine Translation*, PhD thesis, University of Queensland, Brisbane, Australia.

Bond, F. and Vatikiotis-Bateson, C.(2002), Using an ontology to determine English countability, *Proc. of the 19th International Conference on Computational Linguistics (COLING 2002)*, Taipei, Taiwan.

Bouma, G., van Noord, G. and Malouf, R.(2001), Alpino: Wide coverage computational analysis of Dutch, *Computational Linguistics in the Netherlands (CLIN 2000)*, Rodopi, Amsterdam, the Netherlands.

Burnard, L.(2000), *User Reference Guide for the British National Corpus, Technical report*, Oxford University Computing Services.

Clark, S. and Weir, D.(2002), Class-based probability estimation using a semantic hierarchy, *Computational Linguistics*.

Copestake, A.(1992), *Lexical rules in constraint-based grammar*, PhD thesis, University of Sussex, Brighton.

de Buenaga Rodríguez, M., Gómez Hidalgo, J. and Díaz Agudo, B.(2000), Using WordNet to complement training information in text categorization, *in* N. Nicolov and R. Mitkov (eds), *Recent Advances in Natural Language Processing II: Selected Papers from RANLP'97*, Vol. Vol. 189 of *Current Issues in Linguistic Theory (CILT)*, John Benjamins: Amsterdam/Philadelphia.

Fellbaum, C. (ed.)(1998), *WordNet: An Electronic Lexical Database*, MIT Press.

Gillon, B. S.(1996), The lexical semantics of English count and mass nouns, *Proc. of the ACL-SIGLEX Workshop on the Breadth and Depth of Semantic Lexicons*, Santa Cruz, USA, pp. 51–61.

Grishman, R., Macloed, C. and Myers, A.(1998), *COMLEX Syntax Reference Manual*, NYU. Prometheus Project.

Haeseryn, W. et al. (eds)(1997), *Algemene Nederlandse Spraakkunst*, Nijhoff, Groningen.

Jackendoff, R.(1991), Parts and boundaries, *in* B. Levin and S. Pinker (eds), *Lexical and Conceptual Semantics*, Blackwell Publishers, Cambridge MA and Oxford UK, pp. 1–45.

Mandala, R., Tokunaga, T. and Tanaka, H.(2000), Query expansion using heterogeneous thesauri, *Information Processing and Management* **36**(3), 361–78.

O'Hara, T., Salay, N., Witbrock, M., Schneider, D., Aldag, B., Bertolo, S., Panton, K., Lehmann, F., Smith, M., Baxter, D., Curtis, J. and Wagner, P.(2003), Inducing criteria for mass noun lexical mappings using the Cyc KB and its extension to WordNet, *Proc. of the Fifth International Workshop on Computational Semantics (IWCS-5)*, Tilburg, the Netherlands.

Preiss, J., Korhonen, A. and Briscoe, T.(2002), Subcategorization acquisition as an evaluation method for WSD, *Proc. of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Canary Islands.

Schwartz, L. O.(2002), *Corpus-based acquisition of head noun countability features*, Master's thesis, Cambridge University, Cambridge, UK.

Stetina, J. and Nagao, M.(1997), Corpus based PP attachment ambiguity resolution with a semantic dictionary, *Proc. of the 5th Annual Workshop on Very Large Corpora*, Hong Kong, pp. 66–80.

Vossen, P. (ed.)(1998), *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, Kluwer Academic Publishers, Dordrecht.

Wierzbicka, A.(1988), *The Semantics of Grammar*, John Benjamin.