Topics in Corpus-Based Dutch Syntax

Beek, Leonoor Johanneke van der

Publication date:
2005

Chapter 5

Countability

This chapter describes how the countability of nouns can be learned automatically from linguistic resources. Countability is an important lexical feature that determines the syntactic contexts in which nouns can occur, more specifically their ability to combine with or without particular determiners and quantifiers. Countability information is important for accurate and efficient parsing, generation and translation, and we also saw that it was crucial for our research on determinerless PPs in chapter 4. Unfortunately, countability information is not generally available. We therefore experiment with methods to acquire countability information automatically. Two types of linguistic resources are used in this chapter to learn countability: raw text corpora that we preprocess with various linguistic analyzers, and EuroWordNet, a lexical semantic network. In the face of sparse data, we augment our resources with English data and perform crosslingual classification.

5.1 Introduction

In this chapter we discuss two methods for the acquisition of lexical properties of nouns from linguistic resources. We investigate to what extent corpus data on the one hand and ontologies on the other hand are sufficient sources of information for classifying nouns according to their lexical properties, and we compare their relative performance.

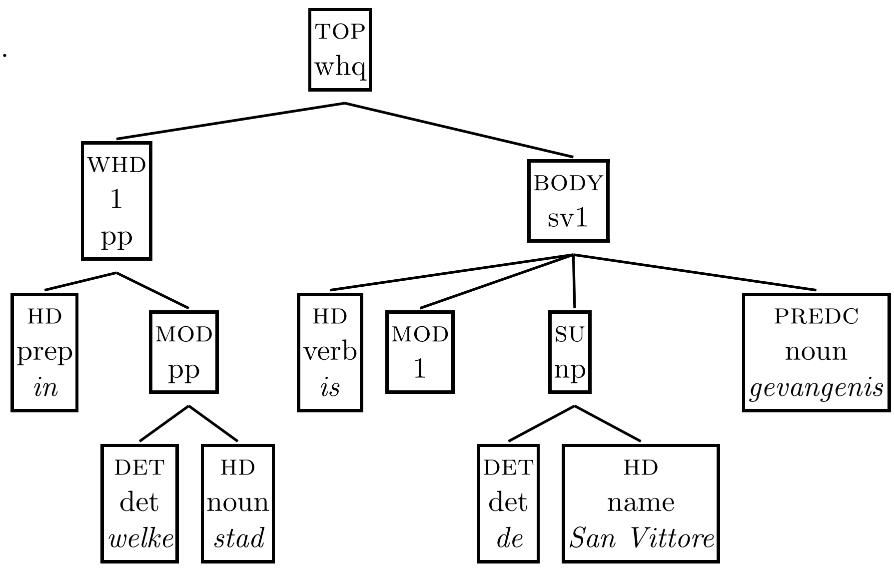
We focus on the linguistic property of countability. Noun countability has not received very much attention in the computational linguistics literature (but see section 5.2 for discussion of some previous work on countability). Nevertheless, it does play an important role in (computational) grammars. The countability of a noun determines its potential to occur with (or without) particular determiners. Singular indefinites form a clear example:

while uncountable indefinite nouns do not combine with a determiner, indefinite countable nouns obligatorily do combine with the indefinite article *een* in Dutch or *a* in English.[1]

Influencing the combinatory potential of nouns, countability is important for language generation and translation. But countability information may also help to reduce the (false) ambiguity of sentences in automatic parsing. We illustrate this point with two examples. The sentences in (1) and (2) are sentences from the Alpino Treebank and the TwNC newspaper corpus. In both cases, the Alpino grammar produced both the correct parses in in (1-c) and (2-c) and the false parses in (1-b) and (2-b). In the first example, the complex NP *San Vittore gevangenis* 'San Vittore jail' functions as the subject of the sentence, as illustrated in (1-c). However, the parser mistakingly splits the complex NP into two separate NPs: a subject NP 'San Vittore' and a predicative complement (PREDC) 'jail' (1-b). This incorrect parse was even considered the best parse. Only once the system knows that the word *gevangenis* 'jail' is countable in Dutch, will it correctly discard the parses in (1-b) as improbable, as the noun in itself cannot saturate an NP.

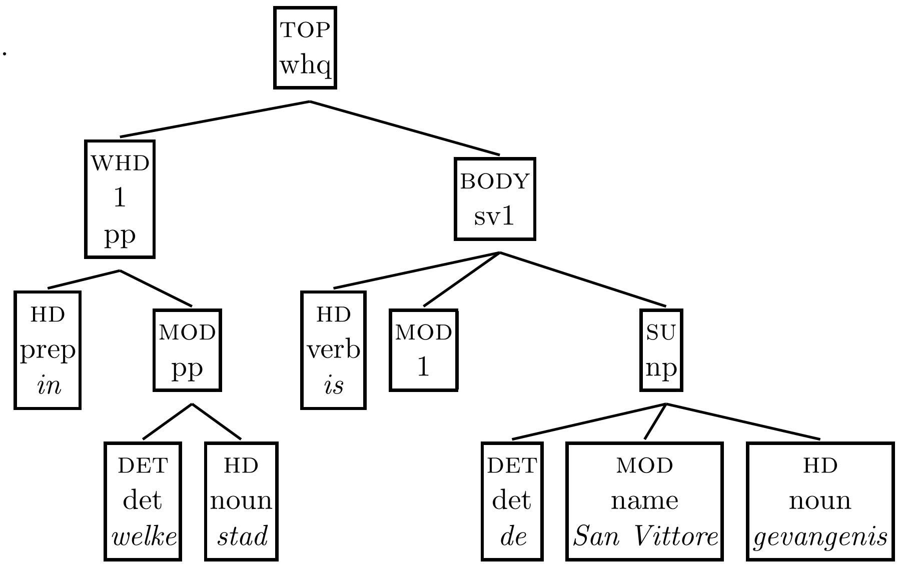
(1) a. In welke stad is de San Vittore gevangenis?
in which city is the San Vittore jail?
In which city is the San Vittore jail?

b.



[1]Except in some predicative constructions, e.g. *Ik wil matroos worden* 'I want to be a sailor' (lit. I want to be sailor).

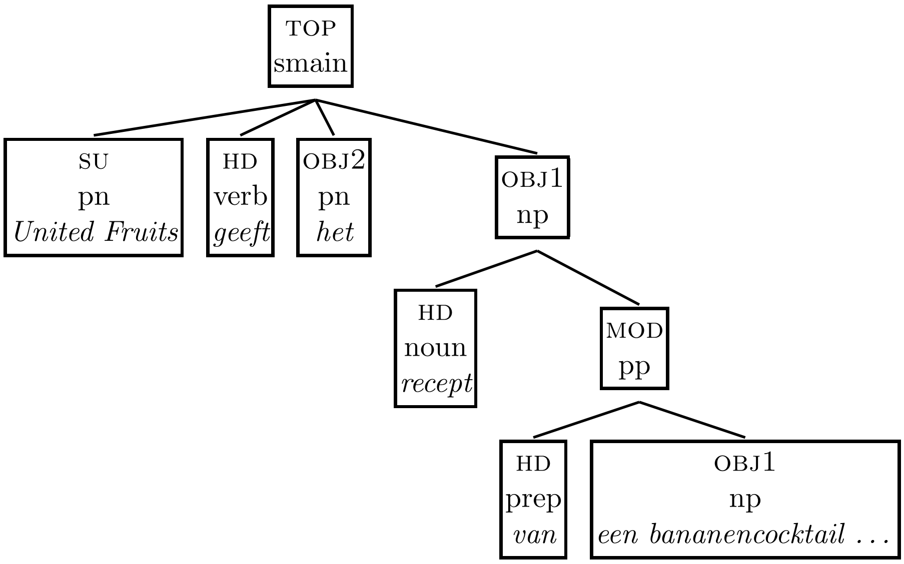
c.



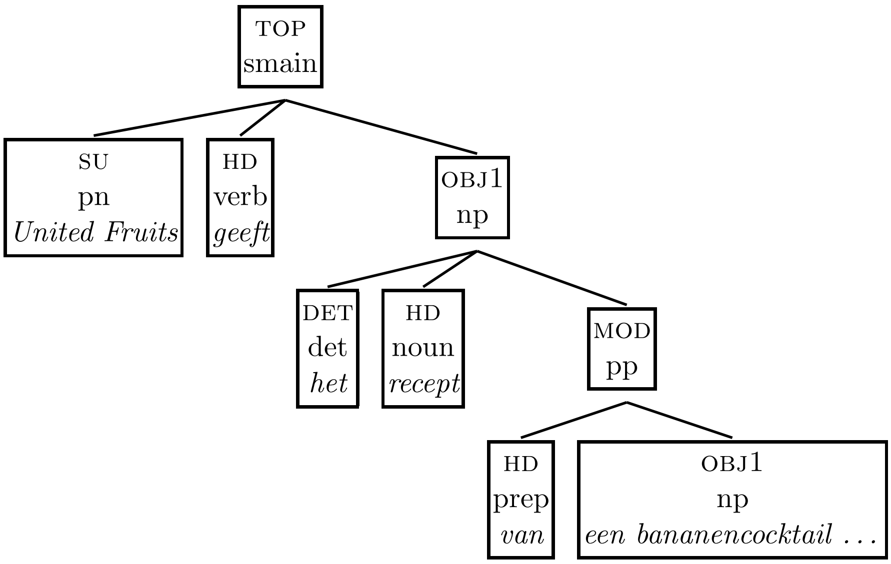
Similarly, the system mistakingly splits up the complex object NP *het recept van een bananencocktail ...* 'the recipe of a banana cocktail' into two NPs. The word *het* is interpreted as the homonymous pronoun *het* 'it' (OBJ2), and analyzed as the indirect object. *Recipe of a banana cocktail* is interpreted as the direct object (OBJ1), resulting in the reading 'United Fruits gives it recipe...'. This incorrect analysis could have been ruled out on the basis of the information that *recipe* is a count noun and thus requires a determiner. Naturally, the ambiguity reduction illustrated in (1) and (2) only works for countable nouns.

(2) a. United Fruit [...] geeft het recept van een bananencocktail,
United Fruits [...] gives the/it recipe of a banana cocktail
die toepasselijk 'Juanita' heet.
that appropriately Juanita is named
United Fruits gives the recipe of a banana cocktail, that is appropriately named 'Juanita'.

b.



c.



Countability may also help word sense disambiguation. Often, two senses of a word have different countabilities. For example, *glas* 'glass' is countable in its drinking equipment sense, but uncountable in its substance reading. Based on this information, we can rule out the drinking equipment reading in (3-a) and the substance reading in (3-b). Only sentence (3-c) remains ambiguous, as the determiner *het* 'the' is compatible with countable and uncountable words.

(3) a. Ik heb glas nodig.
I have glass necessary
I need glass.
b. Ik heb een glas nodig.
I have a glass necessary
I need a glass
c. Ik heb het glas nodig
I have the glass necessary
I need the glass.

Another application of countability information was described in detail in the previous chapter. It was shown that the syntactically-marked determinerless PPs can only be distinguished from the unmarked ones with accurate noun countability information.

We conclude that countability information is crucial for correct and efficient parsing and generation, as well as for the identification of certain syntactically-marked constructions, and that it can resolve certain types of (false) ambiguity. However, countability information is not generally available. In this chapter we try to fill this gap by means of automatic countability classification, experimenting with both corpus-based and ontology-based methods.

The rest of this chapter is structured as follows. In section 5.2 we discuss the notion of countability and the lexical resources that are available for English and Dutch, as well as some earlier work on countability classification. The next two sections discuss the two main approaches to countability classification that we experimented with and the results of both methods: corpus-based countability classification in section 5.3, and ontology-based classification in section 5.4. We end this chapter with a comparison of the various approaches and some concluding remarks in section 5.3.8. The research reported on in this chapter was done in close collaboration with Timothy Baldwin. Parts of this research were previously published in Baldwin and van der Beek (2003) and van der Beek and Baldwin (2004).

5.2 Preliminaries

5.2.1 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.[2] 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 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 deverbalised nouns. Plural-only nouns have only a plural form, and cannot be denumerated: *goederen* 'goods', **drie goederen* '*three goods'. The plural-only class is considered to be a closed class in Dutch. We listed the members of this class in table 5.1. In addition to this list, there are a number of fixed expressions with plurals only nouns, e.g. example (4). As the pluralia tantum are a closed class, the classification experiments below focus exclusively on the countable and uncountable classes, ignoring nouns which are plural only.

(4) Hij zit op zijn hurken.
he sits on his PLURAL ONLY
He is sitting on his heels.

It is important to realize that different senses/usages of a given word can occur with different countabilities, cf. *Ik will 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., *voor mij een rode wijn, graag* 'for me a red wine, please' (countable) vs. *voor mij rode wijn, graag* 'red wine for me, please' (uncountable), which we claim correspond to a single sense of 'wine'.

Accounts of countability range from a purely semantically motivated feature (Jackendoff, 1991) to a completely arbitrary lexical feature in many computational grammars, including the Alpino grammar (Bouma et al., 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.* The Dutch noun is uncountable, whereas the translation in English is countable. An account of countability in terms of a strictly arbitrary lexical feature fails to account for the semantic underpinnings and crosslingual commonalities of countability. Moreover, it implies that type-level countab-

[2]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).

bescheiden	documents	*chemicaliën*	chemicals
conserven	preserves	*contanten*	cash
data	data	*doeleinden*	purpose
echtelieden	marriage partners	*financiën*	finances
gebroeders	brothers	*gegevens*	data
gelieven	lovers	*gemoederen*	minds
genitaliën	genitals	*gezusters*	sisters
goederen	goods	*grutten*	rolled oats
hersenen/hersens	brain(s)	*hurken*	heels
ingewanden	intestines	*inkomsten*	incomings
intimi	friends	*kleren*	clothes
kosten	costs	*levensmiddelen*	provisions
letteren	literature	*manen*	mane
manschappen	manpower	*mazelen*	measles
memoires	memoirs	*mensenrechten*	human rights
middeleeuwen	middle ages	*notulen*	minutes
omstreken	surroundings	*ongeregeldheden*	riots
onkosten	expenses	*onlusten*	riots
paperassen	papers	*papieren*	official documents
personalia	personal data	*troebelen*	disturbances
troepen	troops	*tropen*	tropics
waren	wares	*waterpokken*	chicken-pox
watten	cotton-wool	*zemelen*	bran

Table 5.1: Dutch pluralia tantum.

ility distinctions are categorical, which is in fact not the case. Allan (1980) noted that 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 semantics. This implies that countability is generally stable across languages. But not only countability itself is stable, also the surface effects that noun countability brings about may be very similar for related language pairs. We will see that in both Dutch and English, countability influences the co-occurrence with determiners, certain prepositions and quantity denoting constructions. These two factors, the semantic grounding of countability and the similarities in the effects that countability brings about in different languages, facilitate the crosslingual approaches that we take on countability classification when faced with sparse or medium quality data problems. The crosslingual approach to countability classification taken in section 5.3 crucially relies on the grammatical similarities of countability effects in the aligned languages, while the approach in section 5.4 relies on the semantic basis of countability only.

5.2.2 Lexical resources

For Dutch, few lexical resources with countability information are available. Our Dutch training data consists solely of dictionary data extracted from the Alpino lexicon (Bouma et al., 2001). The Alpino lexicon includes all lexical information found in the Celex electronic dictionary. Countability information was first derived from the presence or absence of a plural form in Celex: all and only *singularia tantum* were considered uncountable. However, the dictionary has been extensively augmented and modified (manually) since. The total number of Dutch nouns is around 14,500. We refer to this set as $\text{Dictionary}_{\text{NL}}$.

In order to test the quality of the Dutch dictionary-derived data and the performance of the classifiers developed in this chapter, we manually annotated 196 unseen Dutch nouns. The nouns were automatically extracted from the POS-tagged NRC part of the Twente Nieuws Corpus.[3] It was decided that the random sample was to be a representative sample of the input of the classifier and should not be edited, leading to an occasional tag-error in the dataset. The countability judgments are based on actual usage in the Twente Nieuws Corpus. Evidence for countability class member ship was extracted from the corpus automatically and checked manually. A noun was classified as a member of each class for which any valid evidence could be found, leading to a very inclusive list. For example, one grammatical example of the plural *gemakken* 'comforts' (e.g. *dat zijn de gemakken van het moderne leven* 'those are the comforts of modern life') would lead to a classification of *gemak* as countable (as well as uncountable). The complete list of nouns and the manually assigned countability judgments can be found in Appendix A. We refer to this dataset as $\text{Annotated}_{\text{NL}}$.

The agreement in countability judgments between $\text{Dictionary}_{\text{NL}}$ and $\text{Annotated}_{\text{NL}}$ is 81.1%. Agreement figures represent the proportion of countability judgments on which both sources agree (i.e. plus or minus countable and plus or minus uncountable for each lexical item). An important part of the disagreement is caused by the fact that nouns in the Alpino dictionary are labeled either countable or uncountable, whereas the nouns in the annotated dataset are potentially labeled as both countable and uncountable.

For English, more data is available. Information about English noun countability was obtained from two lexical sources: COMLEX 3.0 (Grishman et al., 1998) and the common noun part of ALT-J/E's Japanese-to-English semantic transfer dictionary (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 $\text{Dictionary}_{\text{EN}}$ for the remainder of this chapter. In a similar way as for Dutch, 100 unseen nouns were hand-annotated according to actual usage in the British National Corpus (BNC: Burnard (2000)), to make up dataset $\text{Annotated}_{\text{EN}}$. With this dataset, the quality of the English dictionary data was determined. We measured the agreement with $\text{Dictionary}_{\text{EN}}$ to be 85.6%, significantly higher than for the Dutch dictionary data. For an overview of the datasets, see table 5.2.

5.2.3 Past research

Past research on countability classification falls into two basic categories: corpus-based and concept-based.

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

Language	Dataset	Size	Agreement (%)
English (EN)	Dictionary$_{\mathrm{EN}}$	5,853	85.6
	Annotated$_{\mathrm{EN}}$	98	—
Dutch (NL)	Dictionary$_{\mathrm{NL}}$	14,400	81.1
	Annotated$_{\mathrm{NL}}$	196	—

Table 5.2: Countability datasets

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,b) applied this approach to the task of English countability classification in two forms: a) distribution-based classification and b) agreement-based classification. Distribution-based classification is based on the relative frequency of countability related features over token occurrences of a given word. For example, of all occurrences of *rabbit*, how often did it occur without a determiner, or how often did it occur in plural? Distribution-based classification thus looks for feature distribution 'signatures' characteristic of different countabilities. Agreement-based classification looks for convincing evidence of occurrence of one or more features which are uniquely associated with one countability class. For example, the occurrence of a singular noun with the determiner *a* is only possible for countable nouns. The output of multiple pre-processors is used to measure the degree of agreement over the occurrence of those features: an occurrence of a particular feature is used as evidence for a countability class only if multiple preprocessors have observed this fact. Noise introduced by one of the preprocessors is thus filtered out. In evaluation over the four English 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).

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%.

In section 5.3, we attempt to apply the distribution-based methods from Baldwin and Bond (2003a,b) to Dutch. But in contrast to their work, we do not limit ourselves to monolingual classification: we perform crosslingual classification from English to Dutch as a potential solution to the problem of sparse or low quality in-language training data.

Word-to-word countability classification uses direct lexical alignment to determine the countability of novel words from corresponding countability-annotated words. We know of no previous work that applies this strategy, but in section 5.3, we will see that when applied in a crosslingual context using English-to-Dutch word-to-word translation and transliteration data as the source of alignment, the method is remarkably accurate. Transliteration is most accurate, with an accuracy of 98.3%, but has very limited coverage.

Concept-based countability classification, as employed in 5.4, 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 (2002) cite an accuracy of 78% over a 5-way classification of countability preference, whereas O'Hara et al. (2003) 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 the concept-based countability classification in a crosslingual context, as described in section 5.4.

5.3 Corpus-based Classification

The corpus-based approach to countability classification that we take in this section is based on the idea that a noun's countability influences the contexts in which it occurs. It influences for instance the determiners a noun combines with, but also the prepositions and measure nouns that co-occur with it. We perform supervised learning to make a feature 'signature' of each countability class. Nouns are subsequently classified as countable or uncountable based on the contexts in which they occur and which may or may not resemble the signature of a particular class. The performance of the supervised methods is then compared to unsupervised classifiers, which simply look for the occurrence of features which are uniquely associated with one particular countability class.

The supervised classification strategy heavily relies on the quality of the training data. This quality is higher for the English training data than for the Dutch data, with a difference of 4.5% accuracy. Furthermore, English and Dutch are closely related languages, which show the same surface effects of countability. For example, both languages have determiners that combine with one particular countability class, and in both languages uncountable nouns cannot be pluralized or denumerated. Given this similarity and given the fact that better training data is available for English, we decide to experiment with crosslingual classification. The results are compared to the monolingual corpus-based results and to translation-based and transliteration-based crosslingual countability classification.

5.3.1 Feature space

Information about the contexts a noun occurs in is collected in the form of features in a feature space, following Baldwin and Bond (2003a). This feature space is made up of several feature clusters, each of which is conditioned on the occurrence of a target noun in a given construction. The features in those clusters are either one-dimensional or two-dimensional. In the first case, they are simple counts for the occurrence of the target noun in a particular context, for example with the singular determiner *een* 'a'. In the second case, they are counts for the combination of two context factors. An example of a two-dimensional feature is the co-occurrence of the target noun in singular with the number neutral determiner *geen* 'no' or the co-occurrence of the target noun without a determiner and with the preposition *met* 'with'. Below, we provide a basic description of the 9 feature clusters used in this research. After the name of the cluster, we give the number of features in the cluster, both for English (E) and for Dutch (NL). For instance, the twodimensional feature cluster subject-verb agreement is annotated ($^{[2\times2]_{\mathbf{E}}\ vs.\ [2\times2]_{\mathbf{NL}}}$), indicating that on both dimensions (subject number and verb number), there are two realizations possible, resulting in a total of four combinations. Each of those combinations (e.g. singular subject and singular verb) is a twodimensional feature. The value of this feature for a specific noun is the number of times it occurs in singular as the subject of a singular verb. In table 5.3 we list the predicted correlations (table based on Baldwin and Bond (2003a) and adjusted to Dutch).

Head noun number:$^{[2]_{\mathbf{E}}\ vs.\ [2]_{\mathbf{NL}}}$ the number of the target noun when it heads an NP. This captures the fact that countable nouns, but not uncountable nouns, have a plural form.

Subject–verb agreement:$^{[2\times2]_{\mathbf{E}}\ vs.\ [2\times2]_{\mathbf{NL}}}$ the number of the target noun in a subject position vs. number agreement on the governing verb. Another check for plural occurrences of a noun, indicating that it is countable.

Coordinate noun number:$^{[2\times2]_{\mathbf{E}}\ vs.\ [2\times2]_{\mathbf{NL}}}$ the number of the target noun vs. the number of the other noun of the conjunct. This feature is based on the assumption that while most coordinations consist of two plural or two singular conjuncts, uncountable (singular) nouns conjoin with plural nouns more frequently than countable singulars, e.g. *hoofdpijn en tranende ogen* 'headache and burning eyes', *wapens en munitie* 'arms and ammunition'. This is a gradual difference, as countable singulars are by no means impossible with plurals.

N_1 *of* N_2/measure noun constructions: $^{[11\times 2]_{\mathbf{E}}\ \textit{vs.}\ [11\times 2]_{\mathbf{NL}}}$ the number of the target noun (N_2) vs. the type of the N_1 in an English N_1 *of* N_2 construction (e.g. *a group of people*) or Dutch measure noun construction (e.g. *een groep mensen*). We have identified a total of 11 N_1 types for use in this feature cluster (e.g. `COLLECTIVE`, `LACK`, `TEMPORAL`). This captures the fact that measure nouns put restrictions on the countability and the number feature of their complement, e.g. *een kilo appels* 'one kilo of apples' vs. *een kilo suiker* 'one kilo of sugar'.

Occurrence in PPs:$^{[52\times 2]_{\mathbf{E}}\ \textit{vs.}\ [84\times 2]_{\mathbf{NL}}}$ the presence or absence of a determiner when the target noun occurs in singular form in a PP. The purpose of this feature is twofold: occurrence in PPs needs to be treated separate from other occurrences, because of the possibility of determinerless PPs, which does not necessarily indicate that a noun is uncountable (see chapter 4). Furthermore, some prepositions select for countable (*per* 'per') or uncountable (*vol* 'full of') complements.

Pronoun co-occurrence:$^{[12\times 2]_{\mathbf{E}}\ \textit{vs.}\ [7\times 2]_{\mathbf{NL}}}$ what personal, reflexive and possessive pronouns occur in the same sentence as singular and plural instances of the target noun. This features aims at capturing pronoun binding effects: uncountable nouns are not expected to bind a plural pronoun.

Singular determiners:$^{[10]_{\mathbf{E}}\ \textit{vs.}\ [10]_{\mathbf{NL}}}$ what singular-selecting determiners occur in NPs headed by the target noun in singular form. In Dutch, these select singular count nouns (*ieder kind* 'every child' vs. **iedere suiker* '*every sugar'). In English, the determiners may also select for uncountable nouns (e.g. *much sugar*).

Plural determiners:$^{[12]_{\mathbf{E}}\ \textit{vs.}\ [13]_{\mathbf{NL}}}$ what plural-selecting determiners occur in NPs headed by the target noun in plural form. These determiners are not expected to occur with uncountable nouns (*enkele dagen* 'a few days' vs. **enkele tijd* '*a few time').

Number-neutral determiners:$^{[11\times 2]_{\mathbf{E}}\ \textit{vs.}\ [13\times 2]_{\mathbf{NL}}}$ what number-neutral determiners occur in NPs headed by the target noun, and what is the number of the target noun for each. This captures the fact that these determiners combine with plural nouns if the noun is countable, but with a singular noun if it is uncountable (*minder vrije dagen* 'less days off' vs. *minder zout* 'less salt').

Feature cluster	**Countable**	**Uncountable**
Head noun number	S, P	S
Subj-V Agreement	S, P	S
Coordinate noun number	[S,S],[P.P],[P,S]	[S,S],[S,P]
Measure nouns	[*een kilo* 'a kilo of',P],...	[*een kilo* 'a kilo of',S],...
PPs	[*per* 'per',S],...	[*vol* 'full of',S],...
Pronoun co-occurrence	[*hun* 'them',P],...	[*het* 'it',S],...
Singular determiners	[*ieder* 'every',S],...	-
Plural determiners	[*enkele* 'a few',P],...	-
Number-neutral determiners	[*minder* 'less',P],...	[*minder* 'less',S],...

Table 5.3: Predicted values for each feature cluster (S=singular, P=plural)

The Dutch and English feature clusters represent the same linguistic structures, even if the individual features are not direct translations of each other. That is, in both English and Dutch, there are determiners that select for plural (countable) nominals, and in both languages the subject and the verb agree in number. An exception is the Dutch measure noun construction (5-a). In English, the same concept (some quantity of something) is expressed with a different linguistic construction, namely with the N_1 *of* N_2 construction (5-b). The two bring about the same restrictions with respect to countability (5) and thus can be aligned.[4]

(5) a. Een kilo suiker.
   a kilo sugar
 b. A kilo of sugar.
 c. *Een kilo auto.
   a kilo car
 d. *A kilo of car.

5.3.2 Methodology

We use a variety of pre-processors to map the raw data onto the types of constructions targeted in the feature clusters, namely a POS-tagger and a full-text chunker for both Dutch and English, and additionally a dependency parser for English. For Dutch, POS-tags, lemmata and chunk data were extracted from automatically generated, fully parsed Alpino output (Bouma et al., 2001). For English, we used a custom-built fnTBL-based tagger (Ngai and Florian, 2001) with the Penn tagset, morph (Minnen et al., 2001) as our lemmatiser, an fnTBL-based chunker which runs over the output of the tagger, and RASP (Briscoe and Carroll, 2002) as the dependency parser.

These data sets are then used independently to test the efficacy of the different systems at capturing features used in the classification process, or in tandem to consolidate the strengths of the individual methods and reduce system-specific idiosyncrasies in the feature values. When combining the Dutch and English in classification, we invariably combine like systems (e.g. Dutch tagger-derived data with English tagger-derived data).

The Dutch data was extracted from the newspaper (NRC, 13M words)

[4] In fact, the term 'measure noun construction' is an umbrella term for singular, plural and number neutral 'measure nouns', similar to the distinction between singular, plural and number neutral determiners and similar also to the N_1 *of* N_2 construction in English. Although most of these are some sort of measure, we also included nouns like *type* (*een bepaald type auto* 'a certain type of car')

component of the Twente Nieuws Corpus[5] and the English data comes from the written component (90M words) of the British National Corpus (Burnard, 2000).

After generating the different feature vectors for each noun based on the above configurations, we filtered out all nouns which did not occur at least 10 times in NP head position according to the output of all pre-processors. This resulted in 20,530 English nouns and 12,734 Dutch nouns in the training data.

We propose a variety of both monolingual (Dutch-to-Dutch = NN) and crosslingual (English-to-Dutch = EN) unsupervised and supervised classifier architectures for the task of learning countability. We employ two basic classifier architectures: (1) a separate binary classifier for each countability class (**BIN**), and (2) a single multiclass classifier (**MULTI**). The multiclass classifier assigns each noun to one of the three classes 'countable', 'uncountable' or 'both'. A classification in the category 'both' corresponds to a positive classification in both binary classifiers.

In all cases, our supervised classifiers are built using TiMBL version 4.2 (Daelemans et al., 2002), a memory-based classification system based on the k-nearest neighbour algorithm. TiMBL was used with the default configuration except that k was set to 9 throughout.

5.3.3 Monolingual classifiers: design

The various different monolingual classifiers determine the countability of Dutch target nouns on the basis of in-language training material. This training material consist of the 14,400 noun Alpino dictionary data (Dictionary$_{\text{NL}}$), for which we saw that the agreement with the hand-annotated data set was 81.1%. In this section, we discuss the binary classifiers. The multiclass classifiers (both monolingual and crosslingual) are discussed in section 5.3.7.

Unsupervised classifiers

The simplest baseline classifier simply maps all nouns to the most frequent class, which in our case is +countable and −uncountable. In addition to this, we derive a separate baseline for each countability class/pre-processor system combination. We built a (binary) monolingual unsupervised classifier based on diagnostic evidence. For each target noun, the unsupervised classifier simply checks for the existence of diagnostic data in the output of the POS tagger and chunker for the given countability class. Diagnostic data

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

takes the form of unit features which are uniquely associated with a given countability class, e.g. the determiner *een* 'a' co-occurring with a given (singular) noun is a strong indicator of that noun being countable. We refer to these classifiers as $\mathrm{NN}_{\mathrm{BIN}}$(evidence,POS) and $\mathrm{NN}_{\mathrm{BIN}}$(evidence,chunk). We perform basic system combination by voting between the two pre-processor datasets as to whether the target noun belongs to a given countability class, and breaking ties in favour of the POS tagger (NN(evidence,all)).

Distribution-based classifiers: $NN_{\mathrm{BIN}}(feature_{\mathrm{ALL}})$

We implemented a conventional monolingual classifier based on the full feature set given above (section 5.3.1). For each target noun, we compare its value for each feature with the values of other nouns on that feature and the value of the target noun on other features within the feature cluster.

As the absolute frequency of a particular feature-value combination of a noun cannot be compared with the values for other nouns or features, we follow Baldwin and Bond (2003b) in translating each one-dimensional feature f_s for target noun w into three separate feature values, representing the frequency relative to corpus size, word frequency and feature cluster frequency. By means of illustration, we calculate the relative feature values for the feature HEAD NOUN NUMBER$_{sg}$ for word w, which occurred 389 times in singular and 2 times in plural in a 10M word corpus. Suppose that 13 occurrences of the singular noun were in N-N compounds, and in all other (376+2) occurrences the noun was heading an NP. First, we capture the frequency relative to the corpus size:

$$corpfreq(f_s, w) = \frac{freq(f_s|w)}{freq(*)} \tag{5.1}$$

where $freq(*)$ is the frequency of all words in the corpus. For w, this results in: $corpfreq(\textsc{hnn}_{sg}, w) = 376/1{,}000{,}000 = 0.000376$. We furthermore calculate the frequency relative to the target word's frequency:

$$wordfreq(f_s, w) = \frac{freq(f_s|w)}{freq(w)} \tag{5.2}$$

Continuing our example w, we get $wordfreq(\textsc{hnn}_{sg}, w) = 376/391 = 0.962$. The third relative frequency compares our count to the frequencies of the other features in the feature cluster:

$$featfreq(f_s, w) = \frac{freq(f_s|w)}{\sum_i freq(f_i|w)} \tag{5.3}$$

The third relative frequency for our example w is then $featfreq(\text{HNN}_{sg}, w) = 376/378 = 0.995$. Instead of our raw frequency of 376, we now have the 3 relative frequencies 0.000376, 0.962 and 0.995.

In addition to mapping individual unit features onto triples, we introduce a triple for each feature cluster as a whole. This triple represents the sum over all member values.

In case a feature is two-dimensional (e.g. the number of the target noun in subject-position vs. the number of the agreeing verb), each feature $f_{s,t}$ for target noun w is translated into the same relative frequencies $corpfreq(f_{s,t}, w)$, $wordfreq(f_{s,t}, w)$ and $featfreq(f_{s,t}, w)$ as above. In addition, two feature values are introduced which represent the *featfreq* values relative to the totals of each of the two feature dimensions i and j (in combination with the target word). In other words: we calculate the frequency of target word w occurring as the singular subject of a singular verb relative to the frequency of a singular subject w with any verb (singular or plural).

$$featdimfreq_1(f_{s,t}, w) = \frac{freq(f_{s,t}|w)}{\sum_i freq(f_{i,t}|w)} \tag{5.4}$$

$$featdimfreq_2(f_{s,t}, w) = \frac{freq(f_{s,t}|w)}{\sum_j freq(f_{s,j}|w)} \tag{5.5}$$

Finally, we calculate the feature values for the cluster totals for the two-dimensional features. Where this total was a simple sum over all individual feature values for the one-dimensional feature clusters, we now calculate row and column totals. For instance, we calculate totals for each preposition (irrespective of the presence of a determiner) and for determinerless and with-determiner contexts. Each of these totals is described in the form of 3 values, similar to the individual feature values. This methodology is described in Baldwin and Bond (2003b). Given the feature space in section 5.3.1, we generate a total of 1,664 independent values for each target noun.

From the binary Alpino data, individual countable and uncountable classifiers were learned ($\text{NN}_{\text{BIN}}(\text{feature}_{\text{ALL}})$).The feature values in each case were averaged across those from the tagger and chunker.[6]

[6]We additionally built separate classifiers based on the outputs of the individual preprocessors, but found their performance to be inferior to that of the classifier based on their amalgamated output.

Classifier	Acc (%)
Baseline	74.3
NN_{BIN}(evidence,POS)	55.1
NN_{BIN}(evidence,chunk)	50.8
NN_{BIN}(evidence,all)	53.3
NN_{BIN}(feature,all)	81.9
Alpino dictionary	81.1

Table 5.4: Results for monolingual classification

5.3.4 Monolingual classifiers: results and discussion

We compare the overall performance of the different monolingual classifiers. Classifier performance is rated according to accuracy (Acc), i.e. the proportion of correct classifications (table 5.4). For each lexical item, two binary classifications are performed: it is plus or minus countable and plus or minus uncountable. Note that the classifier architecture allows for lexical items to be classified as neither countable nor uncountable.[7] We can compare the scores relative to each other and against a simple baseline. This baseline is a majority class classifier which naively classifies all instances as belonging to the largest class (i.e. +countable and −uncountable).

The most striking result that the unsupervised methods, which were supposed to provide an additional baseline for each combination of countability class and preprocessor, in fact perform considerably worse than our simple majority class baseline.

We zoom in on the unsupervised classification results to see if we can say more about the types of mistakes the classifiers make. We investigate the results for the classification of countable and uncountable noun separately. The results for uncountable nouns (table 5.6) are more accurate, even though the baseline for uncountable (63.8%) is much lower than for countable (84.7%). However, the accuracy remains below the baseline on both countability classes. The precision and recall scores show large differences between countable and uncountable nouns. The tables 5.5 and 5.6 show that irrespective of the preprocessing, precision (P) of the countable classifiers was high. However, this high precision was matched with a low recall (R). For the uncountable classification, we find the reverse situation: a high recall,

[7] In an engineering context, one would use the baseline classifier as a fall-back, mapping the 'unclassified' items to the majority class. The results presented here are based on the system as is. Interestingly, the full feature-based classifiers only failed to classify nouns that can be considered noise in the testset, caused by tag-errors.

Method	**Acc (%)**	**P**	**R**	**F**
NN_{BIN}(evidence,all)	55.1	.964	.488	.648
NN_{BIN}(evidence,chunk)	50.5	.973	.434	.600
NN_{BIN}(evidence,POS)	47.4	.970	.392	.558

Table 5.5: Unsupervised classification results for countable nouns

Method	**Acc (%)**	**P**	**R**	**F**
NN_{BIN}(evidence,all)	55.5	.423	.930	.581
NN_{BIN}(evidence,chunk)	51.0	.414	.887	.565
NN_{BIN}(evidence,POS)	63.8	.490	.718	.583

Table 5.6: Unsupervised classification results for uncountable nouns

but very low precision. In other words: the diagnostics to identify countable nouns are very accurate, but cannot be found very often. The diagnostics for uncountable nouns are more frequently found, but are not very accurate.

These findings are in line with the idea that nouns have a basic countability classification, but may be converted to another countability class: a single occurrence of a diagnostic for uncountable nouns does not mean the noun has a base classification 'uncountable'.

These results may be improved slightly by tweaking the set of diagnostics. For example, only base prepositions were considered as diagnostics, while some collocational prepositions were also shown to pose countability restrictions on their complements in chapter 4 (e.g. *bij wijze van* 'by means of' selects for a countable noun). Nevertheless, the method is not expected to produce reliable lexical information, even with modifications.

The results furthermore show that the unsupervised classifiers that use POS-tagged data only outperform both the chunk-based classifier and the combination of both types of data.

The feature-based classifiers perform much better than the unsupervised classifiers: not only do they outperform the baseline, but with an accuracy of 81.9%, the corpus-based classifiers are also more accurate than the accuracy of the Alpino dictionary training data (81.1%). We expect that there is room for further improvement. As the supervised methods heavily depend on the quantity and quality of the training data, the results may be improved by training on more or better data. Unfortunately, no more or better training material is available for Dutch. For English, on the other hand, high quality dictionary data is available. Although the size of the dataset is smaller than the Dutch dataset (almost 6K words versus more than 14K for Dutch), the quality is higher, with a 85.6% agreement with the hand-annotated dataset, versus 81.1% for Dutch. In addition to this high quality dictionary data, there are large quantities of automatically classified nouns available. Given the fact that English and Dutch are closely related languages, we decide to experiment with crosslingual classification with the higher quality English training data.

5.3.5 Crosslingual classifiers

Below, we describe two ways in which the corpus-based countability classifier can be adjusted to classify Dutch nouns based on English training data. The resulting classifier is compared to its monolingual counterpart and to two other crosslingual approaches: translation-based and transliteration-based classification. We start with a description of each of the crosslingual classifiers.

Translation-based classifier: *EN*$_{\mathrm{BIN}}$ *(translate)*

Translation-based classification applies the observation that Dutch nouns often take the same countability as their English translation equivalents. For this task we use the English automatically classified dataset, which is the output of a monolingual supervised English countability classifier (Baldwin and Bond, 2003a,b). We then extract translation pairs from a bilingual dictionary (English–Dutch freedict version 1.1-1, containing 15,426 Dutch entries) and for each countability class, vote for the membership of a given Dutch noun based on the countabilities of the English translations. In the case that no translation data exists for a given Dutch noun or no countability data exists for the English translations, we classify the Dutch noun countability as 'unknown'. Additionally, we map plural only and bipartite nouns in English onto the Dutch uncountable class.[8]

Transliteration-based classifier: *EN*$_{\mathrm{BIN}}$ *(transliterate)*

Transliteration-based classification relies on the fact that some proportion of the Dutch nouns are spelled the same as their English translations, e.g. *tank, pupil, norm, item, restaurant*. As in translation-based classification, it applies the observation that countability is frequently preserved under translation from English to Dutch, even though some mismatches exist (e.g.

[8]This approach was chosen because the restrictions of plural only and bipartite nouns resemble those of uncountable nouns better than those of countable nouns. In hindsight, mapping of bipartite to countable may have been a better choice, as most translations of bipartite nouns are in fact countable in Dutch.

tissue, which only has a countable sense in Dutch, but has both countable and uncountable uses in English). It takes a Dutch noun and simply determines if a countability-annotated word of the same spelling exists in English, and if so, transfers the countability directly across to Dutch. In all other respects, we implement the method identically to translation-based classification. The advantage of transliteration over translation is that it is resource free. The obvious disadvantage is the expected low coverage.

Cluster-to-cluster classifier: $\boldsymbol{EN}_{\mathrm{BIN}}$ *(cluster)*

As observed above (section 5.3), there is a strong correlation between the feature clusters used for Dutch and English. For example, co-occurrence with plural determiners is a strong indicator that the given noun is countable in both English and Dutch. At the same time, there is generally low correlation between individual unit features. For example, the English plural determiner *many* has no direct Dutch equivalent, and conversely, the Dutch plural determiner *sommige* has no direct English equivalent. The most straightforward way of aligning feature clusters, therefore, is through the (three) amalgamated totals for each one-dimensional feature cluster and some subset of the column and row totals for each two-dimensional feature cluster (e.g. for the PP feature, we align the totals for the singular and plural features but not the totals for each individual preposition independent of number). All values for the individual unit features are then ignored. In this way, it is possible to align 88 feature values, based on the output of the English and Dutch POS taggers.[9] Note that as part of the feature alignment, we take the negative log of all corpus frequency (*corpfreq*) values in an attempt to reduce the effects of differing corpus sizes in English and Dutch (about 90M words for English, vs. 13M for Dutch)

Feature-to-feature classifiers: *EN(feature)*

While we stated above that there is generally low correlation between individual unit features in English and Dutch, some unit features are highly correlated crosslingually. One example is the English singular determiner *a* which correlates highly with the Dutch *een*. Here, we can thus simply match the feature values onto one another directly. In other cases, a many-to-many mapping exists between proper subsets of a given feature cluster (e.g. the

[9] All crosslingual feature-based methods were tested over the output of the POS taggers, the chunkers and the combined outputs of the three English and two Dutch pre-processors. Overall, there was very little separating the results, and the simple POS tagger generally produced the most consistent results, so it is these results we present herein.

English determiner pair *each* and *every* correlates highly with the Dutch determiner pair *ieder* and *elk*), and alignment takes the form of feature value amalgamation in each language by averaging over the unit values, followed by alignment of the amalgamated values. A total of 466 unit feature values are amalgamated into 351 feature values, which are then combined with the 88 aligned total values from cluster-to-cluster classification for a total of 439 feature values. As for cluster-to-cluster classification, we evaluate feature-to-feature classification over the output of the English and Dutch POS taggers.

We implemented a total of 5 feature-to-feature classifiers. The first, $EN_{\mathrm{BIN}}(feature_{\mathrm{ALL}})$, makes use of all aligned features in the form of separate binary classifiers. The second, $EN_{\mathrm{MULTI}}(feature_{\mathrm{ALL}})$, similarly uses all aligned features, but in a multiclass classifier architecture. The other three make use only of a subset of all features: $EN_{\mathrm{BIN}}(feature_{\mathrm{DET}})$ is based on only aligned determiner features, plus the aligned cluster totals; $EN_{\mathrm{BIN}}(feature_{\mathrm{PREP}})$ is based on only aligned preposition features, plus the aligned cluster totals; and $EN_{\mathrm{BIN}}(feature_{\mathrm{PRON}})$ is based on only aligned pronoun features, plus the aligned cluster totals.[10]

System combination: $EN_{\mathrm{BIN}}(combined)$

System combination takes the outputs of heterogeneous classifiers and makes a consolidated classification based upon them. It has been shown to be effective in tasks ranging from word sense disambiguation to tagging in consolidating the performance of component systems (Klein et al., 2002; van Halteren et al., 2001). In our case, we take the outputs of all unsupervised (i.e. evidence-based) and crosslingual classifiers—a total of 12 classifiers—for each countability class, and run TiMBL over them (effectively weighing the influence of each classifier). The 196 Dutch annotated nouns were used as training words for this procedure, and the results are thus based on 10-fold cross-validation. This provides an estimate of the classification performance we could expect over unannotated Dutch noun data using the 196 annotated nouns as training data. Finally, we experimented with a combination of the twelve classifiers above and the Alpino-trained Dutch supervised (binary) classifier.

5.3.6 Crosslingual classification: results and discussion

For a first overview of the results, we compare the accuracy (Acc) and the coverage (Cov) of the various different classifiers on the hand-annotated test

[10]Results for the multiclass classifier over feature subsets were found to be markedly worse than for binary classifiers.

Classifier	**Acc (%)**	**Cov (%)**
Baseline	74.3	100
$\mathrm{NN}_{\mathrm{BIN}}(\mathrm{ALL})$	81.9	100
$\mathrm{EN}_{\mathrm{BIN}}$(transliterate)	98.3	30
$\mathrm{EN}_{\mathrm{BIN}}$(translate)	88.6	36
$\mathrm{EN}_{\mathrm{BIN}}(\mathrm{cluster}_{\mathrm{ALL}})$	77.3	100
$\mathrm{EN}_{\mathrm{BIN}}(\mathrm{feature}_{\mathrm{ALL}})$	72.5	100
$\mathrm{EN}_{\mathrm{BIN}}$(combined)	83.2	100
$\mathrm{E/NN}_{\mathrm{BIN}}$(combined)	85.7	100

Table 5.7: Results for crosslingual classification

Feature	**Acc (%)**
Det	76.0
Prep	76.5
Pron	71.2
All	72.4

Table 5.8: Accuracy for various features.

set in table 5.7. The results for the baseline and the best monolingual (binary) classifiers are included for reference.

The accuracy of the simple translation and transliteration-based classifiers are surprisingly high. Their use is limited, however, because of the low recall. Where al other classifiers always give a positive or negative classification, the translation or transliteration-based classifiers can only classify if a translation or transliteration is available. Failing this, the classifier returns 'unknown'. That is, assuming we have countability data for an English word of the same spelling as a given Dutch noun (or for a translation), we get a very accurate estimate of the Dutch countability.

The crosslingual feature and cluster-based classifiers each individually performed better than baseline, but worse than the monolingual feature-based classifier. The cluster-based classification outperforms the feature-based classification, indicating that important information is lost in the (incomplete) alignment of features. Combined with the high overhead in hand-aligning features in feature-to-feature classification, it is clear that cluster-to-cluster classification should be prefered over feature-based classification.

We investigated the effect of the various individual feature clusters. Table 5.8 list the accuracy of each cluster on the test set. The determiner cluster and the preposition cluster individually perform better than baseline, confirming the assumption that determiner and preposition co-occurrence are influenced by noun countability. However, the pronoun cluster performs worse than baseline and the combination of all clusters leads to a decrease in accuracy. We conclude that pronoun information does not contribute to the correct classification of Dutch nouns as countable or uncountable, even if it appeared helpful for English (Baldwin and Bond, 2003a). This is not to say that pronoun binding information is not influenced by noun countability, but that the proxy of this information in the present study is not a good modeling of pronoun binding information.

System combination proved helpful for countability classification. The combination of crosslingual classifiers (translation, transliteration and distribution-based) performed better than any of the component classifiers. Manually taking out the pronoun cluster (which was shown not to contribute to correct classification) did not change the results at all: running TiMBL over the outputs of all classifiers apparently (correctly) causes the pronoun cluster to receive minimal weight.

The combined crosslingual classifier outperformed monolingual classification. This confirms our claim that given the lack of reliable training data in Dutch, crosslingual classification using English data is a viable option. This finding is particularly striking given that the volume of Dutch training data is more than twice the volume of English data. Having said this, the combined crosslingual/monolingual classifier (EN/NN_{BIN}(combined)) outperforms both the combined crosslingual classifier and the monolingual classifier, in which sense the Alpino data has some empirical utility. That is, we have shown that high-quality out-of-language English countability data is a stronger predictor of Dutch countability than medium-quality in-language Dutch countability data, but at the same time that the two are complementary.

Finally, it should be noted that the classifiers perform much better for countable than for uncountable nouns. To illustrate this effect, we put together the accuracies on countable and uncountable nouns for various classifiers (see table 5.9). This is mainly due to the fact that there is a large difference in the relative occurrence of members of the two classes. Countable nouns are much more frequent than uncountable nouns, resulting in a much higher baseline for countable nouns. For illustration: 84.7% of the nouns in the gold standard data were annotated as countable, vs. 36.2% uncountable (20.9% of the nouns were annotated as having both countable and uncountable uses).

Method	Countable	Uncountable
Baseline	84.7	63.8
$\mathrm{EN_{BIN}}$(translate)	94.8	58.3
$\mathrm{EN_{BIN}}$(transliterate)	100.0	96.6
$\mathrm{EN_{BIN}}$(cluster)	80.6	74.0
$\mathrm{EN_{BIN}}$(feature$_{\mathrm{ALL}}$)	75.0	69.9
$\mathrm{EN_{BIN}}$(combined)	86.2	79.1
EN/$\mathrm{NN_{BIN}}$(combined)	88.8	78.1

Table 5.9: Accuracy for countable and countable nouns

5.3.7 Binary vs. three-way classifiers

For both the monolingual and the crosslingual classifiers, we presented the results of the two binary classifiers: one that classifies nouns as having or not having a countable use and one that classifies nouns as plus or minus uncountable. In addition to these binary classifiers, we implemented three-way classifiers using a selection of the classification methods used for binary classification. The three-way classifiers map the nouns to one of the three classes (strictly) countable, (strictly) uncountable or countable+uncountable. Note that the multiclass classifier does not allow the classification 'none', even though it is possible for a noun to receive a negative classification from both binary classifiers.

The results in table 5.10 are calculated the same way as before: the accuracy is the percentage of correct *classifications*. In other words: a classification as 'both' is counted as +countable, +uncountable, a classification as countable maps to +countable, −uncountable. This way, we can compare the performance of the multiclass classifier with the binary classifier. We see that the results for the threeway classifier are more stable, ranging from 80.6% to 83.4%. It performs better on the worst performing classification methods, but worse on the two best ones, so that the overall best performing classifiers are still the binary crosslingual combined classifier and the binary Dutch+English combined classifier.

5.3.8 Corpus-based approach: conclusion

We have presented several methods for classifying Dutch nouns as countable and/or uncountable on the basis of Dutch and English data. The classifiers depend on translation/transliteration data or linguistic features that were extracted from corpora. We compared a range of crosslingual English-to-

Method	Binary	Multiclass
Baseline	74.2	74.2
NL	81.9	81.4
EN cluster	77.3	80.6
EN feats	72.5	81.6
EN combined	83.2	82.7
EN/NN combined	85.7	83.4

Table 5.10: Accuracy (%) for binary and multiclass classifiers.

Dutch classifiers based on reliable English countability data with monolingual Dutch-to-Dutch classifiers based on lower-quality Dutch countability data, and found that the crosslingual classifiers outperformed the monolingual classifiers to varying degrees. Based on this, we suggest that the optimal fast-track solution to Dutch countability classification is to use English data. We were able to reach a 85.7% agreement with the hand-annotated dataset for the combined crosslingual/monolingual classifier, which is higher than the Alpino dictionary data (81.1%).

We saw that translation and transliteration-based countability classification performed remarkably well given that a translation or transliteration was available. It would be interesting to explore in more detail the possibility of co-training via translation- and transliteration-based classification, as this seems to provide a means for automatically generating high-quality Dutch countability data to learn a monolingual classifier from. The high performance of translation and transliteration-based classification furthermore supports the idea that countability is primarily connected to a semantic concept, rather than to the realization of that concept in a particular language. This hypothesis is further tested in the next section, where we classify nouns on the basis of the countability of its synonyms and other semantically related words both within a language and across languages.

5.4 Ontology-based Classification

Ontologies such as WordNet[11] (Fellbaum, 1998) and EuroWordNet[12] (Vossen and Bloksma, 1998) comprise a hierarchical network of concept nodes, populated with words. The nodes in such networks are conventionally termed 'synsets', as they contain sets of synonymous words representing a com-

[11] http://www.cogsci.princeton.edu/~wn/

[12] http://www.illc.uva.nl/EuroWordNet/

mon 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). In addition, EuroWordNet connects the synsets of various languages, thus facilitating cross-linguistic generalization.

The in-language forms of generalization have been successfully applied in a variety of tasks including text categorization (e.g. de Buenaga Rodríguez et al. (2000)), PP attachment (e.g. Stetina and Nagao (1997)), subcategorization frame acquisition (e.g. Preiss et al. (2002)), selectional preference learning (e.g. Clark and Weir (2002)) and information retrieval (e.g. Mandala et al. (2000)).

This section examines the use of synsets in the automatic acquisition of the countability class 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. There are conflicting 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 et al., 2003).

We take this line of research a step further in exploring the possibilities both for mono- and crosslingual ontology-based countability classification in 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. We thus apply the additional, cross-linguistic, generalization that EuroWordNet facilitates.

We attempt crosslingual classification, as English and Dutch are closely-related languages and the basic nature of noun countability aligns well in the two languages. We already saw in section 5.2.1 that both languages distinguish between the three countability classes of countable, uncountable and plural only,[13] and although mismatches exist—e.g. *hersenen* (plural only) vs.

[13]A fourth class of bipartite nouns (e.g. *scissors, trousers*) is generally recognized for

'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.

In the following, we first outline the lexical resources used in this research, especially where they differ from the resources used in the previous section (section 5.4.1). We then detail the classification procedures (section 5.4.2) and evaluate each method (section 5.4.3).

5.4.1 Lexical resources for WordNet-based classification

For ontology-based classification, we used all the datasets we used for the corpus-based classification, as described in section 5.2.2 and repeated in table 5.11. In addition, we used a second data set consisting of some 11,000 English nouns that were automatically classified on the basis of corpus data (Baldwin and Bond, 2003a,b). In section 5.2.3, we described 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 $\text{Learned}_{\text{EN}}$. The agreement between $\text{Annotated}_{\text{EN}}$ and $\text{Learned}_{\text{EN}}$ is 82.0%, which is still slightly higher than the Dutch dictionary data. In section 5.3 from this chapter, we composed a similar dataset for Dutch. The learned Dutch countability dataset is based on combined monolingual and crosslingual corpus-based and word-to-word classification methods. The methods are applied to around 6,000 common nouns not found in the Alpino lexicon. The agreement for $\text{Learned}_{\text{NL}}$ was higher than for the Dutch dictionary data at a respectable 85.7%, almost identical to that for $\text{Dictionary}_{\text{EN}}$. As with English, we will exclusively use the combination of these datasets in evaluation, which we will refer to as $\textbf{Dic+Learn}_{\text{NL}}$.

Finally, we combined the English dictionary and learned datasets with the corresponding Dutch datasets to form a single multilingual dataset of about 37,000 countability-classified nouns at overall agreement of 83.1%, which we label as $\textbf{Comb}_{\text{EN/NL}}$. For an overview of the datasets, see table 5.11.

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

English, but has no Dutch correlate.

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 5.11: Countability datasets for WordNet-based classification

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 5.11, 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). In addition to the dictionary and the annotated data, we have added the data that was automatically learned in the previous section. Dictionary and learned data are combined in the Dic+Learndatasets. We observe that the $\text{Dic+Learn}_{\text{EN}}$ and the $\text{Dic+Learn}_{\text{NL}}$ datasets are very similar with respect to the number of EWN-mapped words, the agreement with the annotated datasets and the mean level of polysemy.

5.4.2 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 relations we experimented with are (near-)synonyms, hypernyms, hyponyms and cohyponyms. 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.[14] We are thus forced to assign the countability class(es) of each noun to all its senses in EuroWordNet.

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*).

[14] In fact, countabilities in the ALT-J/E lexicon (Bond, 2001) 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.

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).

Cohyponym-based classification Instead of traveling up or down the hierarchy, cohyponym-based classification looks at the countability classes of words that share a hypernym with the target word, i.e. words that are hyponyms of a noun of which the target word is a hyponym, too, a la Bond and Vatikiotis-Bateson (2002). The intuition behind this approach is that although the semantics of those sister synsets may differ considerably, the level of abstraction is the same. We thus increase the amount of training data, without introducing mismatches of the type *tafel* 'table' (countable) vs. its hypernym *meubilair* 'furniture' (uncountable), as in hypernym-based or hyponym-based classification. Instead we compare *tafel* 'table' with *stoel* 'chair', which are both countable. Similar to the other classifiers, the model is cascaded in that it only makes use of the sister information if no (countability classified) synonym is available.

5.4.3 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 5.4.3), and classify using the different EuroWordNet link types (section 5.4.3). We then present a breakdown of the results over countable and uncountable nouns (section 5.4.3), and finally contrast mono- and crosslingual classification (section 5.4.3).

Method	Accuracy (%) Annotated	Accuracy (%) Dic+Learned	Coverage (%) Annotated	Coverage (%) Dic+Learned
Synonyms	74.8	83.3	71.8	70.4
Hyponyms	75.7	78.7	74.5	75.9
Hypernyms	74.8	79.1	97.3	97.0
Hypo+Hyper	73.4	76.1	97.3	97.4
Cohyponyms	76.4	80.6	96.6	95.5

Table 5.12: Accuracy and coverage for various link types (combined classification)

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. If the test set also consists of dictionary and learned data, the results are based on 10-fold cross-validation.

Performance of each EuroWordNet Link Type

The choice of a classification strategy is not independent of the choice for the link types to be included in the training data. Both vary with respect to their degree of strictness: union-based classification is more liberal than majority-based classification and hypernym-based classification is more liberal than synonym-based classification. The stricter the method, the higher the expected accuracy. But we also expect the link types to affect coverage: the more link types are included, the more training words we have and the higher the expected coverage. The classification method does not affect coverage: all methods will classify a target noun if (and only if) at least one training word is found. We therefore start by comparing the accuracy and coverage of different link types.

The results in table 5.12 confirm the intuition that the inclusion of other link types increases coverage. If we restrict ourselves to synonyms, we only find training words for about 70% of the EuroWordNet mapped nouns. Including other link types leads to a 25% increase of this proportion. The results in table 5.12 are for combined classification, but the pattern extends to other classification strategies.

It is interesting to see that the contribution of hyponym data is so much smaller than for hypernym data, even though the increase in training words is at least as big as for hypernyms. An explanation for this fact is that while each synset in EuroWordNet (except for the top node) is associated with (at least) one hypernym synset, many nodes are terminal and thus do not have hyponyms. However, if a synset has a hyponym, it often has many. In other words, many target words have one or a few hypernyms, while few target words have many hyponyms. The effect of this distribution is that including hypernyms leads to many more target words having at least one countability mapped training word, and including hyponyms leads to a few words having many more training words than they had before (if they had any).

Looking at the effect of link type on accuracy, we see that the two data sets differ. While cross-validation on the dictionary and learned datasets shows the expected drop in accuracy, classification of the annotated test set becomes *more* accurate when including other link types. A possible explanation for this result is in the nature of the test set. It is a small test set, and the nouns were randomly selected from (POS-tagged) corpora. As a result, it contains English words (*off, sense*), archaic casemarkings (*state* 'state'), nouns that are almost exclusively used in idiomatic expressions (*toom* 'bridle') and other non-typical nominals. That would then also explain other cases we will see later on where the results on the two test sets differ greatly. However, we would expect these not to be contained in EuroWordNet, so that they only influence coverage,[15] not accuracy.

Another explanation can be found in the number of training words per target word. Compare for example the average of 4 training words per target word in synonym-based classification with the averages of 24, 22, 42 and 10 for hyponyms, hypernyms, both and cohyponyms (training data: Dic+Learn$_{\text{NL}}$, test data: Annotated$_{\text{NL}}$). While the chance of mismatches (*knife* vs. *cutlery*) increases with the introduction of more link types, classification strategies that use voting (i.e. majority-based and combined classification) may benefit from the increase in the average number of training words, as noise may be filtered out. We do not expect the larger average number of training words to increase the accuracy of union-based classification. Indeed, for union-based classification we get a synonym-based score of 76.2% on the annotated test set, vs. 74.3, 70.7 and 73.6% for hyponyms, hypernyms and cohyponyms respectively. It is unclear, however, why this effect is not found when cross-validating over the larger dictionary+learned data set.

In the remainder of this chapter, we focus on the results from cohyponym-based classification. Using hypernym, hypernym+hyponym or cohyponym relations in addition to synonym relations leads to a large increase in cov-

[15] That is: overall coverage, and not coverage relative to the EuroWordNet mapped nouns.

Method	Accuracy (%) Annotated	Dic+Learned
Baseline	74.2	74.2
Union	73.6	73.7
Majority	75.3	81.9
Combined	76.4	80.6

Table 5.13: Cohyponym-based accuracy for different classification methods.

Classification	Union	Majority	Combined
True positive	33	4	14
True negative	57	91	84
False positive	36	2	9
False negative	18	47	37

Table 5.14: Exact cohypononym-based classification results for uncountables.

erage. The effect on accuracy varies depending on the dataset and the classification strategy, but cohyponyms consistently outperform the hypernym and hypernym+hyponym datasets with respect to accuracy.

Performance of each Classification Method

We next investigate the performance of different classification methods (table 5.13). We see that union-based classification performs below baseline. Apparently, the cohyponym data introduces noise, which the union-based classification could not filter out. This is caused by the fact that all evidence for a particular countability class directly leads to a positive classification, even if for a noun n there is only one word pointing to countability class A and a hundred words pointing to countability class B. This is also illustrated in table 5.14 which contains the exact counts for the annotated dataset: the union-based setup too easily classifies positively, resulting in high numbers of true positives, but low numbers of true negatives.

We combined classification for the remainder of the experiments. There is very little separating majority-based from combined classification, but the latter better models the intuition that countability is stable for a given synset. Furthermore, combined classification leads to the highest results overall.

Class	Baseline	Accuracy (%) Annotated	Dic+Learned
Countable	84.7	84.7	82.3
Uncountable	63.8	68.1	78.9
Total	74.2	76.4	80.6

Table 5.15: Accuracy for countable and uncountable classification.

Performance over Countable and Uncountable Nouns

So far, we have averaged our results over the classification of countable nouns and uncountable nouns. However, we saw in section 5.3.6 that there were important differences between the two tasks. Most importantly, the baselines differ greatly: of our 196 item hand-annotated test set, 166 nouns were countable, whereas only 71 were uncountable (41 nouns had both countable and uncountable uses). A majority class baseline classifiers for countables thus performs with an accuracy of 84.7% while the corresponding classifier for uncountables performs with an accuracy of 63.8%.

The different baselines are reflected in the results of ontology-based classification in the same way as we saw for corpus-based classification. Table 5.15 shows the accuracy for uncountable and countable classification separately. Countable classification consistently outperforms uncountable classification. But it is striking that any gain in performance over the baseline comes from uncountable classification. In fact, a combination of a baseline classifier for countables and the cohyponym-based combined classifier for uncountables would yield the highest overall accuracy on the annotated test set, with 82%.

Mono- vs. Crosslingual Classification

The application of an ontology for countability classification was in part motivated by the fact that countability proved stable across related languages such as English and Dutch. The multilingual ontology EuroWordNet provides us with links between synsets in the Dutch network and those in the English network. This means that we can include countability mapped English words to our training set and thus expand the total amount of training data. In table 5.16, it is shown that combining Dutch and English training data leads to an increase in coverage.

More surprisingly, using English training data (with or without the Dutch data) also leads to in increase in accuracy, at least for the annotated test set. That is, crosslingual training data are a more reliable source of information

	Training data					
	Dic+Learn$_{NL}$		**Dic+Learn$_{EN}$**		**Dutch+English**	
Test set	**Acc**	**Cov**	**Acc**	**Cov**	**Acc**	**Cov**
Annotated$_{NL}$	76.4	96.6	80.4	92.6	80.3	98.7
Dic+Learn$_{NL}$	80.6	95.5	80.2	93.2		
Annotated$_{EN}$	82.9	58.6	75.0	51.4	81.2	68.6
Dic+Learn$_{EN}$	80.9	64.7	82.7	54.5		
Dic+Learn$_{EN/NL}$					81.4	84.3

Table 5.16: Accuracy and coverage for mono- and crosslingual classification.

than the in-language data! For the much larger dictionary+learned test set, training on English nouns exclusively leads to a 1% drop in accuracy.

We also performed the reverse classification, testing on English nouns and training on English, Dutch or combined training data. Again, we find the surprising effect that crosslingual classification is more accurate than in-language classification of the hand-annotated testset, with 82.9% for Dutch training data and only 75.0% for English training data. And again, this pattern is reversed if we test on the (English) dictionary+learned dataset. Note that the English annotated test set is even smaller than its Dutch counterpart, with only 70 nouns mapped to EuroWordNet. Finally, cross-validating over the combined Dutch+English dictionary+learned dataset gives an accuracy of 81.4%, vs. 80.3 and 81.2% on the Dutch and English annotated datasets.

We conclude with a word of caution: the annotated datasets are relatively small, and any result must therefore be interpreted with care. Having said this, 17 results are highly suggestive of the finding that using crosslingual training data has little effect on the accuracy of countability classification. Complementing in-language data with crosslingual data furthermore has a positive effect on coverage. It is therefore a successful strategy to increase the performance of classification.

Above, we discarded synonym-based classification, even if it slightly outperformed other link types with respect to accuracy, because of low coverage. But we just found that it is possible to increase coverage by complementing the training set with English data. While this gain in coverage has a modest effect on cohyponym-based classification, which has a high coverage anyway, it may improve synonym-based classification significantly. One might wonder whether synonym-based classification is still outperformed by cohyponym-based classification if more trainingdata is available? Table 5.17 shows that on the annotated dataset, cohyponym-based classification still

	Synonym		Cohyponym	
Test set	**Acc**	**Cov**	**Acc**	**Cov**
Annotated$_{\text{NL}}$	79.4	91.3	80.3	98.7
Dic+Learn$_{\text{EN/NL}}$	83.5	76.7	81.4	84.3

Table 5.17: Accuracy and coverage of synonyms and cohyponyms on the Dutch+English dataset

has both a higher accuracy and a higher coverage. Cross-validation on the larger dataset shows a slight drop in accuracy, but much larger coverage.

If we add the majority classifier to the system as a fallback strategy, words for which no evidence can be found in the ontology will be classified +countable and −uncountable. Both synonym and cohyponym-based classifiers now have a coverage of 100%. On the annotated dataset, cohyponym-based classification outperforms synonym-based classification (79.2% vs. 78.2% accuracy), but the results for cross-validation on the dictionary+learned dataset are the other way around, with 81.1% accuracy for synonyms, and 79.8% for cohyponyms.

5.4.4 Ontology-based classification: conclusion

We have presented several methods for applying EuroWordNet to 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. The methods were applied both to in-language and cross-language 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. Combining Dutch and English data gave the best results, with an accuracy of 80.3% for the Dutch data and 81.2% for the English annotated data.

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.

A major drawback of ontology-based classification is the restriction that the target word must be mapped to EuroWordNet. This was the case for only 149 (76%) of the annotated nouns and 10698 (54%) of the dictionary+learned nouns. This means that even with a coverage of 100% on the EWN-mapped nouns, we will not be able to classify more than 76% and 54% of all nouns in the datasets.

5.5 Conclusion

We investigated two general methods for noun countability classification, each based on one general assumption about countability. First, we noted that a noun's countability influences its potential to combine with particular determiners, measure nouns and so on. We used these differences to compose a 'signature' of syntactic contexts for each countability class, based on the corpus distribution of our training data. Target nouns were classified based on the similarity between this signature and the distribution of the target noun itself.

The second method for noun countability classification is based on the assumption that countability is stable for a given semantics, independent of its realization(s) in a particular language. EuroWordNet was applied to propagate the countability of training words to semantically related target nouns.

Both methods were used for monolingual, crosslingual and combined classification, motivated by limited in-language training data and in both cases crosslingual classification proved a viable solution to (high quality) data sparseness, performing at least as good as in-language classification. Combining mono and crosslingual classification data led to further improvements, outperforming monolingual classification and crosslingual classification independently of the classification strategy.

Of the two general approaches, the corpus-based method proved most successful. We were able to reach 85.7% accuracy on the 196 word hand-annotated test set. The ontology-based approach reached a maximum accuracy of 80.3%. On top of this, the domain of ontology-based classification is restricted to nouns that are mapped to EuroWordNet, whereas the corpus-based method can be applied to any noun occurring in the training corpus.

On the other hand, the potential for application of these methods to other language pairs, is greater for ontology-based classification than for corpus-based classification, as crosslingual corpus-based classification relies on the similarity of the two languages with respect to the surface indicators of countability, whereas ontology-based classification can in principle be carried out for any two languages for which a common ontology exists.