Die Morphologie (f) : Targeted Lexical Acquisition for Languages other than English

Jeremy Nicholson†, Timothy Baldwin†, and Phil Blunsom‡

†‡ Department of Computer Science and Software Engineering
University of Melbourne, VIC 3010, Australia

and

† NICTA Victoria Research Laboratories
University of Melbourne, VIC 3010, Australia

{jeremymn,tim,pcbl}@csse.unimelb.edu.au

Abstract

We examine standard deep lexical acquisition features in automatically predicting the gender of noun types and tokens by bootstrapping from a small annotated corpus. Using a knowledge-poor approach to simulate prediction in unseen languages, we observe results comparable to morphological analysers trained specifically on our target languages of German and French. These results describe further scope in analysing other properties in languages displaying a more challenging morphosyntax, in order to create language resources in a language-independent manner.

1 Introduction

As a result of incremental annotation efforts and advances in algorithm design and statistical modelling, deep language resources (DLRs, i.e. language resources with a high level of linguistic sophistication) are increasingly being applied to mainstream NLP applications. Examples include analysis of semantic similarity through ontologies such as WordNet (Fellbaum, 1998) or VerbOcean (Chklovski and Pantel, 2004), parsing with precision grammars such as the DELPH-IN (Oepen et al., 2002) or PARGRAM (Butt et al., 2002) grammars, and modelling with richly annotated treebanks such as PropBank (Palmer et al., 2005) or CCGbank (Hockenmaier and Steedman, 2002).

Unfortunately, the increasing complexity of these language resources and the desire for broad coverage has meant that their traditionally manual mode of creation, development and maintenance has become infeasibly labour-intensive. As

a consequence, deep lexical acquisition (DLA) has been proposed as a means of automatically learning deep linguistic representations to expand the coverage of DLRs (Baldwin, 2005). DLA research can be divided into two main categories: targeted DLA, in which lexemes are classified according to a given lexical property (e.g. noun countability, or subcategorisation properties); and generalised DLA, in which lexemes are classified according to the full range of lexical properties captured in a given DLR (e.g. the full range of lexical relations in a lexical ontology, or the system of lexical types in an HPSG).

As we attest in Section 2, most work in deep lexical acquisition has focussed on the English language. This can be explained in part by the ready availability of targeted language resources like the ones mentioned above, as well as secondary resources — such as corpora, part-of-speech taggers, chunkers, and so on — with which to aid prediction of the lexical property in question. One obvious question, therefore, is whether the techniques used to perform lexical acquisition in English generalise readily to other languages, where subtle but important differences in morphosyntax might obfuscate the surface cues used for prediction.

In this work, we will examine the targeted prediction of the gender of noun types and tokens in context, for both German and French. As an example, the following phrases display adjectival gender inflection in two of the three languages below.

the open window
das offene Fenster
la fenêtre ouverte

window has no gender in English, *Fenster* is neuter in German and *fenêtre* is feminine in French. So

in English, neither the determiner *the* or the adjective *open* inflect, whereas *das* and *la* are the neuter and feminine forms respectively of the determiner, and *offen* and *ouvert* take the neuter and feminine respective suffixes of *-e*.

On the face of it, the task is remarkably simple: native speakers can achieve near-perfection, and even the accuracy of simplistic morphological analysers is taken for granted. However, both of these rely on significant knowledge of the inflectional morphosyntax of the language, whereas we will take a knowledge-poor approach, and bootstrap from an annotated corpus. An additional motivation for automating gender learning is that we are interested in semi-automated precision grammar development over the full spectrum of languages, from the highest to the lowest density. Given that there is no guarantee we will be able to access a native speaker for a low–density language, automation is the natural course to take. Even if we do have access to a native speaker, we would like to maximise use of their time, and free them up from annotation tasks which we can hope to perform reliably through automatic means.

The knowledge-poor approach is an interesting one — although the features used for prediction are linguistically motivated, we remain agnostic toward a specific target language. Since no language-specific features are being used, the knowledge-poor approach is presumed to generalise over unseen languages, as long as there is consistent, well-defined morphosyntax within it.

Despite its seeming ease, and our examination of gender as a "black-box" learning feature, having gender information is extrinsically valuable in many contexts. For example, natural language generation and machine translation both rely heavily on knowing the gender of a word for accurate inflectional agreement.

The structure of the remainder of this paper is as follows. Section 2 provides a background for deep lexical acquisition and gender prediction. Section 3 describes the language resources of which we made use, and Section 4 details the feature set. Finally, we evaluate our method in Section 5, and supply a discussion and brief conclusion in Sections 6 and 7.

2 Background

2.1 Deep Lexical Acquisition

As mentioned above, DLA traditionally takes two forms: targeted toward a specific lexical property, or generalised to map a term to an amalgam of properties defined for a given resource.

The latter technique is often construed as a classification task where the classes are the lexical categories from the target resource. One example is extending an ontology such as WordNet (e.g. Pantel (2005), Daudé et al. (2000)). Another is learning the categories for the lexicon of a precision grammar, such as the English Resource Grammar (ERG; Flickinger (2002);Copestake and Flickinger (2000)), as seen in Baldwin (2005). A common tool for this is supertagging, where the classes are predicted in a task analogous to part-of-speech (POS) tagging (Clark, 2002; Blunsom and Baldwin, 2006).

The former technique is exemplified by expert systems, where the target language is occasionally not English. These learn properties such as verb subcategorisation frames (e.g. Korhonen (2002) for English or Schulte im Walde (2003) for German) or countability (e.g. Baldwin and Bond (2003) for English, or van der Beek and Baldwin (2004) for Dutch, with a crosslingual component).

Common methodologies vary from mining lexical items from a lexical resource directly (e.g. Sanfilippo and Poznanski (1992) for a machine–readable dictionary), learning a particular property from a resource to apply it to a lexical type system (e.g. Carroll and Fang (2004) for verb subcategorisation frames), restricting possible target types according to evidence, and unifying to a consolidated entry (e.g. Fouvry (2003) for precision grammar lexical types), or applying the lexical category of similar instances, based on some notion of similarity (e.g. Baldwin (2005), also for lexical types). It is this last approach that we use in this work.

Implicit in all of these methods is a notion of the secondary language resource (LR). While the primary LR is the (actual or presumed) resource whose types are targeted by the DLA, a secondary LR is an available resource that can be used to aid acquisition. Common examples, as mentioned above, are corpora, POS taggers, and chunkers. Secondary LRs of varying degrees of complexity are available for some languages; however we examine primarily simple LRs in order to remain

faithful to lower–density languages.

2.2 Gender Prediction

Gender is a morphosemantic phenomenon observed in many Indo-European languages. It is observed generally in three classes: masculine, feminine, or neuter (French has masculine and feminine only). Whereas the only English words that inflect for gender are pronouns, in most Indo-European languages at least nouns, adjectives, and determiners also inflect. This normally occurs by way of suffixation, but some languages, such as Swahili, use prefixation.

Gender appears to be a purely semantic property which is determined based on the underlying shape, manner, or sex of the referent. However, morphology can play a strong role, by way of gender selection according to the morphemes of the wordform. A classic example is *Mädchen* "girl" in German, which is neuter because words with the *-chen* suffix are neuter, despite the obvious feminine semantic connotations of this instance.

The contextual and morphological effects shown in the "open window" example above have been theorised as priming the gender predictions of language users when confronted with unseen words.[1] When contextual or lexicographic information is available for a language, this is usually a reliable method for the prediction of gender. Consequently, automatic prediction of gender in languages which have inflectional morphology is usually seen as the domain of the POS tagger (such as Hajič and Hladká (1998)), or morphological analyser (e.g. GERTWOL (Haapalainen and Majorin, 1995) for German and FLEMM (Namer, 2000) for French).

One work in automatic gender prediction that is similar to this one is the bootstrapping approach of Cucerzan and Yarowsky (2003). Starting with a seed set of nouns whose gender is presumably language–invariant, they mine contextual features to hypothesise the gender of novel instances. They then extract simple morphological features of their larger predicted set, and use these to predict the gender of all nouns in their corpora.

The major differences between this work and our own are in the approach Cucerzan and Yarowsky use, and the classes that they can handle. First, their semi-automatic approach relies on a bilingual dictionary from which to extract the seeds — if a machine readable one does not exist, they annotate the seeds by hand. Our approach is fully automatic and can act with an arbitrary set of seeds (although an arbitrarily pathological set of seeds would perform arbitrarily poorly). Second, their method is only well-defined for predicting gender in languages with only masculine and feminine, as they do not propose canonical neuter noun candidates. Our approach makes no claims on the number or underlying semantics of genders in a language, and can equally be extended to predict other morphosyntactic properties such as case and number, where canonical forms are poorly defined.

3 Secondary Language Resources

We used a number of secondary language resources: most notably annotated and unannotated corpora, as well as inflectional lexicons and a POS tagger.

3.1 Corpora

Our primary data sources were two corpora: the TIGER treebank[2] (Brants et al., 2002) for German and the BAF corpus[3] (Simard, 1998) for French.

TIGER is a corpus of about 900K tokens of German newspaper text from the Frankfurt Rundschau, semi-automatically annotated for lemma, morphology, POS and syntactic structure.

The BAF corpus is a bilingual French–English collection of eleven documents comprising Canadian government proceedings, machine translation technical documents, and a Jules Verne novella. There are about 450K sentence-aligned French tokens, with no annotation of morphology or syntax. This corpus is heavily domain–specific, and the lack of annotation provides particular problems, which we explain below. Note that we make no use of the English component of BAF in this paper.

3.2 Inflectional Lexicons

Whereas our German corpus has gold-standard judgements of gender for each token, the French corpus has no such information. Consequently, we use a semi-automatic method to match genders to nouns. Using the Le*fff* syntactic lexi-

---

[1] See Tucker et al. (1977), among others, for detailed studies of L1 and L2 gender acquisition.

[2] http://www.ims.uni-stuttgart.de/projekte/TIGER
[3] http://rali.iro.umontreal.ca/Ressources/BAF

con[4] (Sagot et al., 2006) and Morphalou[5] (Romary et al., 2004), a lexicon of inflected forms, we automatically annotate tokens for which the sources predict an unambiguous gender, and hand-annotate ambiguous tokens using contextual information. These ambiguous tokens are generally animate nouns like *collègue*, which are masculine or feminine according to their referent, or polysemous nouns like *aide*, whose gender depends on the applicable sense.

3.3 POS Taggers

Again, TIGER comes annotated with hand-corrected part-of-speech tags, while the BAF does not. For consistency, we tag both corpora with TreeTagger[6] (Schmid, 1994), a decision tree–based probabilistic tagger trained on both German and French text. We were interested in the impact of the accuracy of the tagger compared to the corrected judgements in the corpus as an extrinsic evaluation of tagger performance. The tagger token accuracy with respect to the TIGER judgements was about 96%, with many of the confusion pairs being common nouns for proper nouns (as the uniform capitalisation makes it less predictable).

4 Deep Lexical Acquisition

We use a deep lexical acquisitional approach similar to Baldwin (2005) to predict a lexical property. In our case, we predict gender and restrict ourselves to languages other than English.

Baldwin examines the relative performance on predicting the lexical types in the ERG, over various linguistically-motivated features based on morphology, syntax, and an ontology: the so-called "bang for the buck" of language resources.

To take a similar approach, we extract all of the common nouns (labelled NN), from each of the corpora to form both a token and a type data set. We generate our feature set independently over the token data set and the type data set for both morphological and syntactic features (explained below) without feature selection, then perform 10-fold stratified cross-validation using a nearest-neighbour classifier (TiMBL 5.0: Daelemans et al. (2003)) with the default $k = 1$ for evaluation.

A summary of the corpora appears in Table 1.

Corpus	Tokens	NN Tokens	NN Types
TIGER	900K	180K	46K
BAF	450K	110K	8K

Table 1: A summary of the two corpora: TIGER for German and BAF for French. The comparatively low number of noun types in BAF is caused by domain specificity.

4.1 Morphological DLA

Morphology-based deep lexical acquisition is based on the hypothesis that words with a similar morphology (affixes) have similar lexical properties. Using character n-grams is a simple approach that does not require any language resources other than a set of pre-classified words from which to bootstrap.

For each (token or type) instance, we generate all of the uni-, bi-, and tri-grams from the word-form, taken from the left (prefixes and infixes) and the right (suffixes and infixes), padded to the length of the longest word in the data. For example, the 1-grams for *fenêtre* above would be *f, e, n, ê, t, r, e, #, #, ...* from the left (L) and *e, r, t, ê, n, e, f, #, #, ...* from the right (R).

We evaluate using each of 1-, 2-, and 3-grams, as well as the combination of 1- and 2-grams, and 1-, 2-, and 3-grams from L and R and both (LR) — to make $5 \times 3 = 15$ experiments.

Other resources for morphological analysis exist, such as derivational morphological analysers, lemmatisers, and stemmers. We do not include them, or their information where it is available in our corpora, taking the stance that such tools will not be readily available for most languages.

4.2 Syntactic DLA

Syntax-based deep lexical acquisition purports that a lexical property can be predicted from the words which surround it. Most languages have at least local morphosyntax, meaning that morphosyntactic and syntactico-semantic properties are attested in the surrounding words.

For each token, we examine the four word forms to the left and right, the POS tags of these word-forms, and corresponding bi-word and bi-tag features according to (Baldwin, 2005). For each type, we take the best N of each feature across all of the relevant tokens.

We examine both left (preceding, in languages written left–to–right) and right (following) context

to maintain a language agnostic approach. While in general, contextual gender information is encoded in the noun modifiers, it is unclear whether these modifiers precede the noun (as in head–final languages like English or German), follow the noun, or occur in some combination (as in French).

While context in freer–word–order languages such as German is circumspect, limiting our feature set to four words on each side takes into account at least local agreement (which is a fixture in most languages). Other unrelated information can presumably be treated as noise.

A summary of the feature set appears in Table 2.

4.3 Ontological DLA

We do not make use of an ontology in the way that Baldwin does; although a candidate ontology does exist for these particular languages (EuroWord-Net; Vossen (1998)), the likelihood of such as resource existing for an arbitrary language is low.

5 Evaluation

We evaluate each of the feature sets across the four data sets collected from the two corpora: the German NN-tagged tokens in TIGER, the German NN-tagged types in TIGER, the French NN-tagged tokens in BAF and the French NN-tagged types in BAF. There were four lexical types for German: MASC, FEM, NEUT, and * and three lexical types for French: MASC, FEM, *. The instances labelled * were those to which a gender could not be sensibly ascribed (e.g. abbreviations such as *PGs*), or uniquely defined (e.g. gender-underspecified nouns such as *relâche*).

Baseline accuracy for each set corresponds to the majority–class: for German tokens and types, this was FEM (41.4 and 38.4% respectively); for French tokens and types, this was MASC (51.8 and 52.7%).

5.1 Morphology

The results for the morphological features are shown in Table 3, for 1, 2, and 3-grams taken from the left and right.

The performance over tokens is excellent, comparable to that of language-specific morphological analysers. Taking characters from the right unsurprisingly performs best, as both German and French inflect for gender using suffixes.

	German		French	
	token	*type*	*token*	*type*
L1	93.8	77.0	99.4	85.5
L2	93.6	77.0	99.4	88.0
L3	93.4	73.5	99.4	86.0
L1+2	93.8	77.3	**99.5**	87.5
L1+2+3	93.6	75.1	99.4	87.3
R1	97.1	85.3	**99.5**	87.9
R2	**97.4**	**86.6**	**99.5**	**88.4**
R3	96.9	84.2	99.4	85.3
R1+2	**97.4**	86.5	**99.5**	**88.4**
R1+2+3	97.3	85.9	**99.5**	87.5
LR1	95.6	78.5	99.4	85.5
LR2	96.2	82.0	99.4	86.1
LR3	95.7	78.9	99.4	85.7
LR1+2	96.1	81.6	99.4	86.1
LR1+2+3	96.0	80.9	99.4	85.5

Table 3: Morphological results using TiMBL

	German		French	
	token	*type*	*token*	*type*
All	82.2	52.5	—	—
TT POS	81.7	52.1	95.5	66.6
WF only	**84.9**	**53.9**	**96.6**	**67.6**

Table 4: Syntactic results using TiMBL

The best results invariably occurred when using bigrams, or unigrams and bigrams together, suggesting that gender is encoded using more than just the final character in a word. This is intuitive, in that a 1-letter suffix is usually insufficient evidence; for example, *-n* in French could imply a feminine gender for words like *maison, information* and a masculine gender for words like *bâton, écrivain*.

5.2 Syntax

The syntax results shown in Table 4 show the gold-standard POS tags (ALL) against those estimated by TreeTagger (TT POS), when combined with wordform context. We also contrast these with using the wordforms without the part-of-speech features (WF only). For type results, we took the best N features across corresponding tokens — for consistency, we let $N = 1$, i.e. we considered the best contextual feature from all of the tokens.[7]

---
[7]Experimentally, a value of 2 gave the best results, with a constant decrease for larger N representing the addition of ir-

Feature type	Positions/description
MORPHOLOGY	
Left	1-, 2-, 3-, 1+2-, 1+2+3-grams
Right	1-, 2-, 3-, 1+2-, 1+2+3-grams
Left/Right	1-, 2-, 3-, 1+2-, 1+2+3-grams
SYNTAX	
Wordform	$-4, -3, -2, -1, +1, +2, +3, +4$
POS tag	$-4, -3, -2, -1, +1, +2, +3, +4$
Bi-Word	$(-3, -2), (-3, -1), (-2, -1), (+1, +2), (+1, +3), (+2, +3)$
Bi-Tag	$(-4, -1), (-3, -2), (-3, -1), (-2, -1), (+1, +2), (+1, +3), (+1, +4), (+2, +3)$

Table 2: Morphological (n-gram) and syntactic (contextual) features.

Both token-wise and type-wise results are much poorer than the ones observed using morphological features. This is unsurprising, firstly because gender is primarily a morphological feature, and is encoded in syntax only through inflection of contextual wordforms. Also, often contextual evidence for gender is weak — for example, nouns beginning with a vowel in French do not take the canonical *le, la*, only *l'* (e.g. *l'ami* (m)); similarly, plural words in German do not inflect for gender: i.e. instead of taking *der, die, das*, plural nouns only take *die* (e.g. *die Freunde* (m)).

In fact, gender is so much a morphological feature that removing the part-of-speech features uniformly improves results. Again, this is unsurprising, seeing as the contextual parts of speech impact only weakly on gender preferences.[8] We return to discuss POS features below.

6 Discussion

The application of language–inspecific features to the task of gender prediction was quite successful, with both morphological and syntactic features comfortably outperforming both the type and token baselines in both German and French. This, however, is not a stunning achievement, as a rule–based system built by hand in a few minutes by a native speaker of the target language could also boast such claims.

The morphological features, based on character n-grams, performed much better than the syntactic features, based on contextual wordform and part-of-speech features. This is validated by the natural claim that the morphemes and semantic gender of a word are somehow linked more strongly than gender to its syntactic properties. Capturing the clues in adjectival and determiner inflection, shown in various examples above, is more challenging for an automatic system.[9]

We attest that the observed improvement in performance between gender prediction in German and French, especially at the token level, is not an indication of a simpler task in French, only a domain–specific corpus. While TIGER is a treebank of newspaper data, the BAF is a small number of topical texts, with little variability.

This specificity perhaps is best evinced through the ratio of tokens to types: for German there are approximately 4 tokens/type, for French, this number balloons to almost 14, despite the two corpora being of the same relative size. Having a large number of exemplars in the training split will almost certainly bias prediction, as the gold-standard tag is usually known. There is minimal false evidence: the proportion of multi-gendered types is only about 3%.

Consequently, the results over noun types are more interesting than those over noun tokens, as they smooth to some extent the multiple corpus instances and domain effects. For types, morphological features taken from the left are still much better in French than German, but those taken from the right give similar results. Syntactic features are consistently better as well. It would be interesting to contrast these with results taken from a French treebank, to parallelise the results for German.

As mentioned above, using bigrams or a combination of unigrams and bigrams generally gives

---

relevant features. A more generous match over any of the corresponding features might alleviate this problem somewhat.

[8]Contextual parts-of-speech generally are uninformative for gender: consider whether masculine, feminine, or neuter nouns are more likely to be followed by a finite verb. This is not exclusively the case, however, as subcategorisation frames are occasionally influenced by gender (e.g. the propensity of deverbal nouns ending in *-tion* (f) to take a *de-*complement); we saw no evidence of this in our data, however.

[9]Approaches like pertinence (Turney, 2006) using a very large corpus could help to mine contextual features for an unknown language. Of course, the probability of having a very large corpus for an unknown language is low, so the problem is somewhat circular.

the best performance for morphological features. This contrasts with the approach taken by Cucerzan and Yarowsky (2003), who extract unigram suffix morphological features. We hypothesise that having longer features may give this technique better discrimination, although this remains to be seen for other languages.

It is not surprising that suffix morphological features perform better than those based on prefixes for these languages, what is surprising is that the morphological features taken from the left work at all. On a token level, this can be partially explained by instances of exact matches occurring in the training data. On the type level, we surmise that there is enough training data for the classifier to accurately predict gender according to instances of uniform length. This hypothesis is supported by reducing the cross-validation split to 10%-90% (effictively simulating a low-density language); for German, unigrams from the left drop to 56% accuracy, while unigrams from the right only fall to 75%.

While the poor performance of the syntactic features leads us to conclude that they are unreliable for this particular task, they may still be fruitful in extending this approach to other lexical properties. A similar morphosyntactic property to gender is case, but this is more heavily syntactic and a method based on morphology only is likely to struggle.

In retaining our syntactic features, we analyse the performance particularity of the POS tagger. While it has a 4% error rate over tokens, a drop of only a few tenths of a percentage is observed in place of gold-standard tags. With wordform context by itself performing better, having an accurate POS tagger seems an inefficient use of our resources, as it is only moderately available across target languages. However, there are syntactico-semantic properties such as countability and sub-categorisation frames which rely on syntactic distinctions that are almost irrelevant for morphosyntactic phenomena (e.g. particle vs. preposition confusion). The downstream application of simplistic POS taggers remains to be seen for these tasks.

Obvious extensions to this work, as mentioned above, are making use of a French treebank, and examining other morphosyntactic lexical properties such as number and case, or syntactico-semantic properties such as countability and sub-

categorisation frames. Taking several simple morphosyntactic properties into account could lead to a language–independent morphological analyser.

Just as important, however, is an analysis of these types of properties (where they exist) for languages with a markedly different morphosyntax. Examples are complex case systems seen in Eastern European languages, agglutinative morphology such as in Turkish, or infixing as in several Austronesian languages.

Finally, the preponderance of data available in English (among other languages) makes cross–lingual deep lexical acquisition tempting. Similarly to Cucerzan and Yarowsky (2003), where a small bilingual dictionary exists, it seems possible to bootstrap from a high–volume data set to that of a smaller language, presumably by learning the underlying lexical semantics (e.g. the countability learning in van der Beek and Baldwin (2004)). One telling question, however, is the necessary "closeness" of the source and target language for this to be feasible.

7 Conclusion

We presented an analysis of standard deep lexical acquisition features in naively predicting gender automatically for German and French noun tokens and types. The morphological features performed comparably to analysers trained on the target languages, while the syntactic features provide scope for other morphosyntactic lexical features. This methodology could aid in construction of language resources in a language–independent manner.

References

Timothy Baldwin and Francis Bond. 2003. Learning the countability of English nouns from corpus data. In *Proc. of the 41st Annual Meeting of the ACL*, pages 463–470, Sapporo, Japan.

Timothy Baldwin. 2005. Bootstrapping deep lexical resources: Resources for courses. In *Proc. of the ACL-SIGLEX Workshop on Deep Lexical Acquisition*, pages 67–76, Ann Arbor, USA.

Phil Blunsom and Timothy Baldwin. 2006. Multilingual deep lexical acquisition for HPSGs via supertagging. In *Proc. of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, pages 164–171, Sydney, Australia.

Sabine Brants, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, and George Smith. 2002. The TIGER treebank. In *Proc. of the Workshop on Treebanks and Linguistic Theories (TLT02)*, Sozopol, Bulgaria.

Miriam Butt, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. 2002. The parallel grammar project. In *Proc. of the 2002 COLING Workshop on Grammar Engineering and Evaluation*, Taipei, Taiwan.

John Carroll and Alex Fang. 2004. The automatic acquisition of verb subcategorisations and their impact on the performance of an HPSG parser. In *Proc. of the 1st International Joint Conference on Natural Language Processing (IJCNLP-04)*, pages 107–114, Sanya City, China.

Timothy Chklovski and Patrick Pantel. 2004. VerbOcean: Mining the web for fine-grained semantic verb relations. In *Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pages 33–40, Barcelona, Spain.

Stephen Clark. 2002. Supertagging for combinatorial categorial grammar. In *Proc. of the 6th International Workshop on Tree Adjoining Grammars and Related Frameworks (TAG+6)*, pages 19–24, Venice, Italy.

Ann Copestake and Dan Flickinger. 2000. An open-source grammar development environment and broad-coverage english grammar using HPSG. In *Proc. of the Second conference on Language Resources and Evaluation*, Athens, Greece.

Silviu Cucerzan and David Yarowsky. 2003. Minimally supervised induction of grammatical gender. In *Proc. of the 3rd International Conference on Human Language Technology Research and 4th Annual Meeting of the NAACL (HLT-NAACL 2003)*, pages 40–47, Edmonton, Canada.

Walter Daelemans, Jakub Zavrel, Ko van der Sloot, and Antal van den Bosch. 2003. *TiMBL: Tilburg Memory Based Learner, version 5.0, Reference Guide*. ILK Technical Report 03-10.

Jordi Daudé, Lluis Padró, and German Rigau. 2000. Inducing ontological co-occurrence vectors. In *Proc. of the 38th Annual Meeting of the ACL*, pages 125–132, Ann Arbor, USA.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, USA.

Dan Flickinger. 2002. On building a more efficient grammar by exploiting types. In Stephan Oepen, Dan Flickinger, Jun'ichi Tsujii, and Hans Uszkoreit, editors, *Collaborative Language Engineering*. CSLI Publications, Stanford, USA.

Frederik Fouvry. 2003. *Robust Processing for Constraint–Based Grammar Formalisms*. Ph.D. thesis, University of Essex, Colchester, UK.

Mariikka Haapalainen and Ari Majorin. 1995. GERTWOL und morphologische disambiguierung für das Deutsche. In *Proc. of the 10th Nordic Conference on Computational Linguistics*, Helsinki, Finland.

Jan Hajič and Barbora Hladká. 1998. Tagging inflective languages: Prediction of morphological categories for a rich, structured tagset. In *Proc. of the 36th Annual Meeting of the ACL and 17th International Conference on Computational Linguistics (COLING/ACL-98)*, pages 483–490, Montréal, Canada.

Julia Hockenmaier and Mark Steedman. 2002. Acquiring compact lexicalized grammars from a cleaner treebank. In *Proc. of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Spain.

Anna Korhonen. 2002. *Subcategorization Acquisition*. Ph.D. thesis, University of Cambridge, Cambridge, UK.

Fiametta Namer. 2000. FLEMM : un analyseur flexionnel du Français à base de règles. *Traitement Automatique des Langues*, 41:523–548.

Stephan Oepen, Dan Flickinger, Jun'ichi Tsujii, and Hans Uszkoreit, editors. 2002. *Collaborative Language Engineering. A Case Study in Efficient Grammar-Based Processing*. CSLI Publications, Stanford, USA.

Martha Palmer, Paul Kingsbury, and Daniel Gildea. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Patrick Pantel. 2005. Inducing ontological co-occurrence vectors. In *Proc. of the 43rd Annual Meeting of the ACL*, pages 125–132, Ann Arbor, USA.

Laurent Romary, Susanne Salmon-Alt, and Gil Francopoulo. 2004. Standards going concrete: from LMF to Morphalou. In *Proc. of the COLING-2004 Workshop on Enhancing and Using Electronic Dictionaries*, pages 22–28, Geneva, Switzerland.

Benoît Sagot, Lionel Clément, Éric Villemonte de La Clergerie, and Pierre Boullier. 2006. The le*fff* syntactic lexicon for French: Architecture, acquisition, use. In *Proc. of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1348–1351, Genoa, Italy.

Antonio Sanfilippo and Victor Poznanski. 1992. The acquisition of lexical knowledge from combined machine–readable dictionary sources. In *Proc. of the 3rd Conference on Applied Natural Language Processing (ANLP 1992)*, pages 80–87, Trento, Italy.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proc. of the 1st International Conference on New Methods in Language Processing*, Manchester, UK.

Sabine Schulte im Walde. 2003. *Experiments on the Automatic Induction of German Semantic Verb Classes*. Ph.D. thesis, Institut für Maschinelle Sprachverarbeitung, Universität Sttutgart, Stuttgart, Germany.

Michel Simard. 1998. The BAF: A corpus of English and French bitext. In *Proc. of the 1st International Conference on Language Resources and Evaluation (LREC'98)*, Granada, Spain.

G. Richard Tucker, Wallace E. Lambert, and André Rigault. 1977. *The French speaker's skill with grammatical gender: an example of rule-governed behavior*. Mouton, The Hague, Netherlands.

Peter D. Turney. 2006. Expressing implicit semantic relations without supervision. In *Proc. of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL 2006)*, pages 313–320, Sydney, Australia.

Leonoor van der Beek and Timothy Baldwin. 2004. Crosslingual countability classification with EuroWordNet. In *Papers from the 14th Meeting of Computational Linguistics in the Netherlands*, pages 141–155, Antwerp, Belgium.

Piek Vossen, editor. 1998. *EuroWordNet: a multilingual database with lexical semantic networks for European Languages*. Kluwer, Dordrecht, Netherlands.