

Scalable Deep Linguistic Processing: Mind the Lexical Gap

Timothy Baldwin

Department of Computer Science and Software Engineering
and

NICTA Victoria Laboratories

University of Melbourne
VIC 3010, Australia

tim@csse.unimelb.edu.au

Abstract

Coverage has been a constant thorn in the side of deployed deep linguistic processing applications, largely because of the difficulty in constructing, maintaining and domain-tuning the complex lexicons that they rely on. This paper reviews various strands of research on deep lexical acquisition (DLA), i.e. the (semi-)automatic creation of linguistically-rich language resources, particularly from the viewpoint of DLA for precision grammars.

1 Introduction

Over recent years, computational linguistics has benefitted considerably from advances in statistical modelling and machine learning, culminating in methods capable of deeper, more accurate automatic analysis, over a wider range of languages. Implicit in much of this work, however, has been the existence of **deep language resources** (DLR hereafter), that is resources which encode precise symbolic linguistic knowledge. DLRs aim either to capture a particular feature of a language, such as verb argument structure (e.g. COMLEX (Grishman et al., 1994) or PropBank (Palmer et al., 2005)) or lexical semantics (e.g. WordNet (Fellbaum, 1998) or FrameNet (Baker et al., 1998)), or alternatively to model a language in its entirety, in the form of a precision grammar (e.g. the English Resource Grammar (Flickinger, 2002), various ParGram grammars (Butt et al., 2002) or CCGBank (Hockenmaier and Steedman, 2007)).

In line with Baldwin (2005a), we consider the development of DLRs to be made up of two basic tasks: (1) design of a data representation to systematically capture the generalisations and idiosyncracies of the dataset of interest (**system design**); and (2) classification of data items according to the predefined data representation (**data classification**). In the case of a deep grammar, for example, system design encompasses the construction of the system of lexical types, templates, and/or phrase structure rules, and data classification corresponds to the determination of the lexical type(s) each individual lexeme conforms to. Naturally, the two tasks are tightly integrated, and actual DLR construction involves successive iterations over the design of analyses (with the data classification informing the system design), and the application of those analyses over data (with the system design informing the data classification).

One common problem pervading all DLRs is that of coverage, which we define as the proportion of relevant data points in a representative text that are adequately described in a given DLR. Based on our bifurcation of the DLR development process, lack of coverage is caused by either: (a) deficiencies in the system design, e.g. a precision grammar not containing an analysis of a given construction, or a lexical semantic DLR not capturing a particular verb frame type; or (b) deficiencies in data classification, that is certain words simply not having been classified, or being only partially described in the DLR. This arises through simple resource constraints on DLR development, domain effects, or the DLR not capturing generalisations effectively (e.g. not supporting conversion between noun countabilities).

While system design is a highly specialised, manually-intensive task with little scope for automation to battle the effects of coverage,¹ the data classification lends itself readily to automation, particularly once the system design is relatively mature and the DLR in question has been populated with seed instances. The focus of this paper is the task of automated data classification, or **deep lexical acquisition** (DLA), in the context of maximising the lexical coverage of a given DLR.

The purpose of this paper is to give a brief introduction into the current state of DLA, centring around precision grammars. As part of this, we present a classification of DLA research along three axes: general-purpose vs. targeted, *in vitro* vs. *in vivo*, and token- vs. type-based. Note that while the bulk of our examples and references relate to English, all claims about the nature of DLA are intended to be language-inspecific.

DLA research falls into two broad categories: general-purpose DLA and targeted DLA. **General-purpose DLA** is a generic approach to DLA which is applicable to any DLR, e.g. in learning the senses for a given word in a wordnet, or the lexical types (or equivalent) for a given lexical entry in a precision grammar. **Targeted DLA**, on the other hand, takes the form of a dedicated expert system or classifier for a specific lexical feature.

DLA methods generally use preprocessing or feature extraction to model lexical items, providing another basis for classification according to the relationship between the preprocessing/feature extraction method and the target DLR: *in vitro* methods use secondary lexical resource(s) to model lexical items, whereas *in vivo* methods are embedded within the target DLR and extract features directly from it.

Finally, DLA methods can be classified as either type- or token-based. **Type-based** methods make predictions about lexical items out of context, e.g. about whether a given verb can participate in various subcategorisation frames, while **token-based** methods make predictions about a lexical item in a given context, e.g. about the precise subcategorisation associated with a given lexical item in a given context.

In the remainder of this paper, we first describe the recent confluence of DLR development and statistical natural language processing (Section 2). We then briefly outline and review each of general-purpose and targeted DLA (Section 3), *in vitro* and *in vivo* DLA (Section 4), and type- and token-based DLA (Section 5).

2 DLR Development and Statistical NLP

This paper is certainly not novel in proposing that machine learning and statistical natural language processing (NLP) enter the fray of DLR development. While the conventional view is that there exists an unbridgable gulf between DLR development and statistical NLP, in practice an increasingly symbiotic relationship has developed between the two (Baldwin et al., 2007). Indeed, statistical NLP has provided a strong consumer base for DLRs, as illustrated by the “core” English language technologies of:

- POS tagging (developed primarily using the Penn Treebank (Marcus et al., 1993))
- treebank parsing (based almost exclusively on the Penn Treebank (Marcus et al., 1993))
- word sense disambiguation (based on SemCor (Landes et al., 1998) and the Senseval tasks)
- semantic role labelling (based on PropBank (Palmer et al., 2005) and FrameNet (Baker et al., 1998))

That is not to say that it has been a one-way street, however, and modern-day DLR developers are calling upon statistical NLP techniques in more and more of their day to day work. In this paper, we focus specifically on the import of (largely) statistical DLA in the context of DLR expansion, to help scale DLRs to greater coverage.

¹Although there have been notable successes in porting system designs across languages for a given resource type, e.g. in ontology induction (Maedche, 2002; Nichols et al., 2005).

3 Applicability

DLA methods differ in their relative portability to different DLRs, with general-purpose DLA being applicable to any arbitrary DLA task, and targeted DLA being tailored to a specific sub-task.

3.1 General-purpose DLA

General-purpose DLA methods are designed to be applicable to any DLR, and generally employ a combination of type- and token-level features. For example, Baldwin (2005a) uses a set of character-level n -grams based on the lexeme and derivational features from a derivational lexicon, in addition to token-level contextual features from the simple word context, a POS tagger, a chunk tagger and a full parser. Pantel and Pennacchiotti (2006) and Snow et al. (2006) independently proposed methods of automatically inducing templates for use in learning semantic relations, but in a manner which could equally be applied to any other DLA task. In this same vein, Joanis and Stevenson (2003) propose a general-purpose verb feature set which they show to be applicable in a range of DLA tasks relating to English verbs.

For token-level DLA tasks, general-purpose DLA can take the form of supertagging over a sufficiently general feature set. **Supertagging** can be defined as the process of applying a sequential tagger to the task of predicting the lexical categorie(s) associated with each word in an input string, relative to a given DLR. It was first introduced as a means of reducing parser ambiguity by Bangalore and Joshi (1999) in the context of the LTAG formalism, and has since been applied in a similar context within the CCG formalism (Clark and Curran, 2004). Baldwin (2005c) used supertagging to learn token-level lexical types for the ERG, based on the full template set of the FNTBL 1.0 English tagger (Ngai and Florian, 2001). Blunsom and Baldwin (2006) revisited these experiments with a generalised feature set, and performed the lexical type prediction task relative to both the ERG and the JACY grammar of Japanese (Siegel and Bender, 2002).

A third form of general-purpose DLA is resource alignment, which can be as general as simply assuming that the two DLRs (i.e. system designs) correspond to a common tree or graph data structure. One common example of resource alignment is mining lexical items directly from a DLR (e.g. a machine-readable dictionary (Sanfilippo and Poznański, 1992) or WordNet (Daudé et al., 2000)).

3.2 Targeted DLA

With targeted DLA, a specialised methodology is proposed to (automatically) learn a particular linguistic property. The outputs from a suite of targeted DLA methods are then combined to form fully-specified lexical items. For example, in the following lexical entry for the noun *admission* in the English Resource Grammar (Copestake and Flickinger, 2000; Flickinger, 2002):

```
admission_n1 := n_pp_mc-of_le &
[ STEM < "admission" >,
  SYNSEM [ LKEYS.KEYREL.PRED
           "_admission_n_of_rel",
           PHON.ONSET voc ] ].
```

the key values are the lexical type (`n_pp_mc-of_le`) and the onset (`voc`).² The lexical type states that the word is a noun (`n`) which optionally subcategorises for a PP (`pp`), is either countable or uncountable (`mc`), and that the selected PP is headed by *of* (`of`). The onset value states that the word has a vowel (rather than consonant) onset. Given separate targeted DLA classifiers for subcategorisation frames and countability, we could construct the fully-specified lexical type for the lexical item, and combine this with the onset prediction to complete the lexical entry.

Subcategorisation frame learning is one of the most widely-researched areas of targeted DLA (Brent (1993), Manning (1993), Briscoe and Carroll (1997), Korhonen (2002), Schulte im Walde (2003), inter alia). Systems generally employ tagging and/or parsing to form token-level subcategorisation frame hypotheses,

²Note that the third value in the lexical entry, other than the lexeme itself, i.e. ("`a_admission_n_of_rel`"), is a unique string identifier for the predicate associated with the lexical entry, which can be set to an arbitrary value.

based on manually-constructed templates. Various statistical filters are applied to these hypotheses to form type-level predictions.

Corpus-based noun countability learning operates in a similar fashion, in tagging and/or parsing sentences containing a noun of interest, and running extraction templates over the output of the preprocessors. These are then pooled together as inputs to a classifier which returns a type-level prediction (Baldwin and Bond, 2003; Nagata et al., 2006). Similarly to subcategorisation learning, the filtering of the tagger/parser output analysis is highly tuned to countability classification, and not immediately reusable for any other DLA task.

This same general approach can also be used to learn the lexemes themselves in the case of multiword expressions. For example, Baldwin (2005b) learns which verbs combine with intransitive prepositions to form verb particle constructions in English, and at the same time predicts the lexical type(s) of each such verb particle.

Other examples of targeted DLA include ontology alignment (Knight and Luk, 1994; Atserias et al., 1997; Daudé et al., 2000; Klein, 2001), grammatical gender prediction (Cucerzan and Yarowsky, 2003; Nicholson et al., 2006), classifying the semantics of derivational affixes (Light, 1996), and verb aspect learning (Siegel and McKeown, 2000).

Note that it is possible to hybridise targeted and general-purpose DLA, e.g. in building or reusing a targeted DLA system to learn lexical properties in one format and then mapping this onto the DLR of choice via a general-purpose alignment method (e.g. Carroll and Fang (2004) for verb subcategorisation learning).

4 Reliance on Secondary DLRs

DLA systems tend to be made up of a variety of preprocessors, feature extractors and machine learners. To spell out the reliance of a given method on external resources, and its relative separation from the target DLR, we differentiate between *in vitro* and *in vivo* DLA.

4.1 In Vitro DLA

As the name suggests, *in vitro* DLA is based on analysis of lexemes in a context independent of the DLR we are looking to learn lexical items for. That is, we make use of a secondary LR or independent preprocessor to model lexical similarity, and use the target DLR only in classifying training instances.

In vitro DLA can be the only means available of performing DLA if we do not have access to annotated data for a given DLR. This would be the case if we were wanting to carry out DLA over a WordNet-style lexical ontology for which sense-annotated data did not exist,³ or over a precision grammar which did not have sufficient coverage to parse significant amounts of corpus data.

The most widely-practised method of *in vitro* DLA extrapolates away from a DLR to corpus or web data, in analysing occurrences of words in template-based contexts which are predicted to correspond to particular lexical types, e.g. as seen above for targeted DLA (Section 3.2).

In vitro DLA can also take the form of resource translation, as seen above for general-purpose DLA (Section 3.1).

4.2 In Vivo DLA

In vivo DLA directly leverages the target DLR to learn new lexical items. This is most commonly performed via an annotated corpus, e.g. a sense-annotated corpus in the case of a lexical ontology, or parsed data of some description (e.g. a treebank) in the case of a precision grammar. Note that we do not consider raw corpus data to be a secondary LR as long as any filtering/analysis of the data is performed based on techniques derived directly from the target DLR (e.g. using the tokeniser built in to a precision grammar), independent of any external pre-processor. That is, DLA which aligns templates from raw corpus data directly with classes in the target DLR is considered to be *in vivo* DLA.

³Noting the possibility of learning from monosemous relatives, c.f. Leacock et al. (1998) and Martinez et al. (2006).

A particularly poignant example of *in vivo* DLA is the research of Fouvry (2003) on token-based DLA for unification-based precision grammars. Here, partially-specified lexical features are generated for an unknown word in a given sentence context via the constraints of syntactically-interacting words, and combined to form a consolidated lexical entry for that word. That is, rather than relying on indirect feature signatures to perform lexical acquisition, the DLR itself drives the incremental learning process.

Supertagging is also a classic instance of *in vivo* DLA, as all of its feature extraction and optimisation over label sequences is based on the target DLR, with the qualification that we classify a precision grammar and its associated treebank, e.g., as a single DLR due to the tight linkage between the two (Oepen et al., 2002; Hockenmaier and Steedman, 2007).

In an intriguing piece of research, Zhang and Kordoni (2005) combine these two in a two-stage *in vivo* token-level DLA system for unknown words in precision grammars. First, a targeted DLA system is run over each unknown word in a given sentence to predict an *N*-best list of lexical type hypotheses (similarly to a supertagger, but without the sequential learning element). These are passed onto the target precision grammar, which generates a parse forest for the given sentence. The final selection of lexical type is then made by the grammar proper in unpacking the preferred parse from the parse tree, based on the parse selection module.

Once again, it is theoretically possible to hybridise between *in vitro* and *in vivo* DLA, e.g. in modifying the Zhang and Kordoni (2005) method to have the first-stage classifier incorporate features from a POS tagger and/or chunker, making the first step *in vitro*. If these predictions are passed onto the grammar and parse selection module as before, the second step would remain *in vivo*.

5 Data Point Granularity

Finally, DLA methods can either be applied to token-level instances of a given lexical item, or alternatively perform type-level classification.

It is important to note that DLRs can similarly be type-based, such as wordnets or precision grammars, or token-based, such as treebanks or sensebanks. That is not to say, however, that type-based DLRs are exclusively associated with type-based DLA methods, or token-based DLRs with token-based DLA methods. For example, subcategorisation lexicons are type-level, but subcategorisation learning is conventionally performed by performing token-level analysis of the subcategorisation properties of individual token instances of a given lexical item, and combining this into a type-level prediction (see Section 3.2). Equivalently for (token-level) sensebanking—that is, the annotation of lexical items with word sense information in a given context—it is possible to have type-level sense induction inform the sensebanking (Agirre and Soroa, 2007).

5.1 Token-level Classification

We have seen a number of token-level classification tasks in our discussion above, namely: supertagging and token-level unknown word lexical type prediction for precision grammars. Another example is multiword expression identification (Lapata and Lascarides, 2003; Kim and Baldwin, 2006),

5.2 Type-level Classification

Once again, we have discussed a number of type-level DLA tasks above, including: subcategorisation learning, countability learning, type-level unknown word lexical type prediction for precision grammars, grammatical gender prediction, and ontology induction and alignment. Other examples include word sense discrimination (Schütze, 1998; Agirre and Soroa, 2007), semantic relation harvesting (Hearst, 1992; Pantel and Pennacchiotti, 2006; Snow et al., 2006), learning qualia structure (Bouillon et al., 2002; Yamada and Baldwin, 2004), and extraction of multiword expressions (Baldwin, 2005b; Baldwin, 2005d; Villada Moirón, 2005).

6 Conclusion

This paper has presented a broad survey of research on deep lexical acquisition, particularly in the context of precision grammars. We developed a classification of DLA methods along three broad axes: general-purpose vs. targeted, *in vitro* vs. *in vivo*, and token- vs. type-level classification. We hope this will provide a framework for continued cross-fertilisation across tasks and methodologies in the area.

Acknowledgements

This paper builds off research with a number of collaborators, including Phil Blunsom, Francis Bond, Ann Copestake, Dan Flickinger, Valia Kordoni, Jeremy Nicholson, Stephan Oepen and Yi Zhang. I would like to thank the PACLIC-21 organisers for giving me the opportunity to present this research.

References

- Agirre, Eneko and Aitor Soroa. 2007. SemEval-2007 task 02: Evaluating word sense induction and discrimination systems. In *Proc. of the 4th International Workshop on Semantic Evaluations*, pages 7–12, Prague, Czech Republic.
- Atserias, Jordi, Salvador Climent, Xavier Farreres, German Rigau, and Horacio Rodríguez. 1997. Combining multiple methods for the automatic construction of multilingual wordnets. In *Proc. of RANLP 1997 (Recent Advances in Natural Language Processing)*, Tzigrav Chark, Bulgaria.
- Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proc. of the 36th Annual Meeting of the ACL and 17th International Conference on Computational Linguistics: COLING/ACL-98*, pages 86–90, Montreal, Canada.
- Baldwin, Timothy. 2005a. Bootstrapping deep lexical resources: Resources for courses. In *Proc. of the ACL-SIGLEX 2005 Workshop on Deep Lexical Acquisition*, pages 67–76, Ann Arbor, USA.
- Baldwin, Timothy. 2005b. The deep lexical acquisition of English verb-particle constructions. *Computer Speech and Language, Special Issue on Multiword Expressions*, 19(4):398–414.
- Baldwin, Timothy. 2005c. General-purpose lexical acquisition: Procedures, questions and results. In *Proc. of the 6th Meeting of the Pacific Association for Computational Linguistics (PACLING 2005)*, pages 23–32, Tokyo, Japan. (Invited Paper).
- Baldwin, Timothy. 2005d. Looking for prepositional verbs in corpus data. In *Proc. of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pages 180–9, Colchester, UK.
- Baldwin, Timothy and Francis Bond. 2003. Learning the countability of English nouns from corpus data. In *Proc. of the 41st Annual Meeting of the ACL*, pages 463–70, Sapporo, Japan.
- Baldwin, Timothy, Mark Dras, Julia Hockenmaier, Tracy Holloway King, and Gertjan van Noord. 2007. The impact of deep linguistic processing on parsing technology. In *Proc. of the 10th International Workshop on Parsing Technologies (IWPT-2007)*, pages 36–8, Prague, Czech Republic.
- Bangalore, Srinivas and Aravind K. Joshi. 1999. Supertagging: An approach to almost parsing. *Computational Linguistics*, 25(2):237–65.
- Blunsom, Phil 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 164–71, Sydney, Australia.
- Bouillon, Pierrette, Vincent Claveau, Cécile Fabre, and Pascale Sébillot. 2002. Acquisition of qualia elements from corpora – evaluation of a symbolic learning method. In *Proc. of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Canary Islands.
- Brent, Michael R. 1993. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics*, 19(2):243–62.

- Briscoe, Ted and John Carroll. 1997. Automatic extraction of subcategorization from corpora. In *Proc. of the 5th Conference on Applied Natural Language Processing (ANLP)*, pages 356–63, Washington DC, USA.
- Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. 2002. The Parallel Grammar project. In *Proc. of the COLING-2002 Workshop on Grammar Engineering and Evaluation*, pages 1–7, Taipei, Taiwan.
- Carroll, John and Alex Fang. 2004. The automatic acquisition of verb subcategorisations and their impact on the performance of an HPSG parser. In *Proc. of the First International Joint Conference on Natural Language Processing (IJCNLP-04)*, pages 107–14, Sanya City, China.
- Clark, Stephen and James R. Curran. 2004. The importance of supertagging for wide-coverage CCG parsing. In *Proc. of the 20th International Conference on Computational Linguistics (COLING 2004)*, pages 282–8, Geneva, Switzerland.
- Copestake, Ann and Dan Flickinger. 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proc. of the 2nd International Conference on Language Resources and Evaluation (LREC 2000)*, Athens, Greece.
- Cucerzan, Silviu 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 40–7, Edmonton, Canada.
- Daudé, Jordi, Lluís Padró, and German Rigau. 2000. Mapping WordNets using structural information. In *Proc. of the 38th Annual Meeting of the ACL*, Hong Kong, China.
- Fellbaum, Christiane, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, USA.
- Flickinger, Dan. 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.
- Fouvry, Frederik. 2003. *Robust Processing for Constraint-based Grammar Formalisms*. Ph.D. thesis, University of Essex.
- Grishman, Ralph, Catherine Macleod, and Adam Myers. 1994. COMLEX syntax: Building a computational lexicon. In *Proc. of the 15th International Conference on Computational Linguistics (COLING '94)*, pages 268–272, Kyoto, Japan.
- Hearst, Marti. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proc. of the 14th International Conference on Computational Linguistics (COLING '92)*, Nantes, France.
- Hockenmaier, Julia and Mark Steedman. 2007. CCGbank: a corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics*, 33(3):355–96.
- Joanis, Eric and Suzanne Stevenson. 2003. A general feature space for automatic verb classification. In *Proc. of the 10th Conference of the EACL (EACL 2003)*, pages 163–70, Budapest, Hungary.
- Kim, Su Nam and Timothy Baldwin. 2006. Automatic identification of English verb particle constructions using linguistic features. In *Proc. of the Third ACL-SIGSEM Workshop on Prepositions*, pages 65–72, Trento, Italy.
- Klein, Michel. 2001. Combining and relating ontologies: An analysis of problems and solutions. In *Proc. of the IJCAI-2001 Workshop on Ontologies and Information Sharing*, Seattle, USA.
- Knight, Kevin and Steve K. Luk. 1994. Building a large-scale knowledge base for machine translation. In *Proc. of the 12th Annual Conference on Artificial Intelligence (AAAI-94)*, pages 773–8, Seattle, USA.
- Korhonen, Anna. 2002. *Subcategorization Acquisition*. Ph.D. thesis, University of Cambridge.
- Landes, Shari, Claudia Leacock, and Randee I. Tengi. 1998. Building semantic concordances. In Christiane Fellbaum, editor, *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, USA.
- Lapata, Mirella and Alex Lascarides. 2003. Detecting novel compounds: The role of distributional evidence. In *Proc. of the 10th Conference of the EACL (EACL 2003)*, Budapest, Hungary.

- Leacock, Claudia, Martin Chodorow, and George A. Miller. 1998. Using corpus statistics and WordNet relations for sense identification. *Computational Linguistics*, 24(1):147–65.
- Light, Marc. 1996. Morphological cues for lexical semantics. In *Proc. of the 34th Annual Meeting of the ACL*, pages 25–31, Santa Cruz, USA.
- Maedche, Alexander D. 2002. *Ontology Learning for the Semantic Web*. Springer.
- Manning, Christopher D. 1993. Automatic acquisition of a large subcategorization dictionary from corpora. In *Proc. of the 31st Annual Meeting of the ACL*, pages 235–42.
- Marcus, Mitchell P., Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn treebank. *Computational Linguistics*, 19(2):313–30.
- Martinez, David, Eneko Agirre, and Xinglong Wang. 2006. Word relatives in context for word sense disambiguation. In *Proc. of the Australasian Language Technology Workshop 2006*, pages 42–50, Sydney, Australia.
- Nagata, Ryo, Atsuo Kawai, Koichiro Morihiro, and Naoki Isu. 2006. Reinforcing English countability prediction with one countability per discourse property. In *Proc. of COLING/ACL 2006*, pages 595–602, Sydney, Australia.
- Ngai, Grace and Radu Florian. 2001. Transformation-based learning in the fast lane. In *Proc. of the 2nd Annual Meeting of the North American Chapter of Association for Computational Linguistics (NAACL2001)*, pages 40–7, Pittsburgh, USA.
- Nichols, Eric, Francis Bond, and Daniel Flickinger. 2005. Robust ontology acquisition from machine-readable dictionaries. In *Proc. of the 19th International Joint Conference on Artificial Intelligence (IJCAI-2005)*, pages 1111–6, Edinburgh, UK.
- Nicholson, Jeremy, Timothy Baldwin, and Phil Blunsom. 2006. Die morphologie (f): Targeted lexical acquisition for languages other than English. In *Proc. of the Australasian Language Technology Workshop 2006*, pages 67–74, Sydney, Australia.
- Oepen, Stephan, Kristina Toutanova, Stuart Shieber, Christopher Manning, Dan Flickinger, and Thorsten Brants. 2002. The LinGO Redwoods Treebank: Motivation and preliminary applications. In *Proc. of the 19th International Conference on Computational Linguistics (COLING 2002)*, pages 1253–7, Taipei, Taiwan.
- Palmer, Martha, Paul Kingsbury, and Dan Gildea. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.
- Pantel, Patrick and Marco Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proc. of COLING/ACL 2006*, pages 113–20, Sydney, Australia.
- Sanfilippo, Antonio and Victor Poznański. 1992. The acquisition of lexical knowledge from combined machine-readable dictionary sources. In *Proc. of the 3rd Conference on Applied Natural Language Processing (ANLP)*, pages 80–7, Trento, Italy.
- Schulte im Walde, Sabine. 2003. *Experiments on the Automatic Induction of German Semantic Verb Classes*. Ph.D. thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. Published as AIMS Report 9(2).
- Schütze, Hinrich. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.
- Siegel, Eric V. and Kathleen McKeown. 2000. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–627.
- Siegel, Melanie and Emily M. Bender. 2002. Efficient deep processing of Japanese. In *Proc. of the 3rd Workshop on Asian Language Resources and International Standardization*, Taipei, Taiwan.
- Snow, Rion, Daniel Jurafsky, and Andrew Y. Ng. 2006. Semantic taxonomy induction from heterogeneous evidence. In *Proc. of COLING/ACL 2006*, pages 801–8, Sydney, Australia.
- Villada Moirón, Begoña. 2005. *Data-driven identification of fixed expressions and their modifiability*. Ph.D. thesis, Alfa-Informatica, University of Groningen.

- Yamada, Ichiro and Timothy Baldwin. 2004. Automatic discovery of telic and agentive roles from corpus data. In *Proc. of the 18th Pacific Asia Conference on Language, Information and Computation (PACLIC 18)*, pages 115–26, Tokyo, Japan.
- Zhang, Yi and Valia Kordoni. 2005. A statistical approach towards unknown word type prediction for deep grammars. In *Proc. of the Australasian Language Technology Workshop 2005*, pages 24–31, Sydney, Australia.