Special Issue of JETAI on Memory-Based Language Processing
Introduction. JETAI Volume 11 Number 3, Taylor and Francis, 287-292, 1999.

Memory-Based Language Processing. Introduction to the Special Issue.

Walter Daelemans
ILK, Computational Linguistics
Tilburg University
P.O.Box 90153
NL-5000 LE Tilburg, The Netherlands

Abstract

Memory-Based Language Processing (MBLP) views language processing as being based on the direct reuse of previous experience rather than on the use of rules or other structures extracted from that experience. In such a framework, language acquisition is modeled as the storage of examples in memory, and language processing as analogical or similarity-based reasoning. We briefly discuss the properties and origins of this family of techniques, and provide an overview of current approaches and issues.

1 Empirical Natural Language Processing

Natural Language Processing (NLP) studies the knowledge representation and problem solving issues involved in learning, producing, and understanding language. Language Technology, or Language Engineering, uses the formalisms and theories developed within NLP in applications ranging from spelling error correction to machine translation and automatic extraction of knowledge from text.

Although the origins of NLP are both logical and statistical, as in other disciplines of Artificial Intelligence, the knowledge-based approach has historically dominated this field. This has resulted in an emphasis on logical semantics for meaning representation, on the development of grammar formalisms (especially lexicalist unification grammars), and on the design of associated parsing methods and lexical representation and organization methods. Well-known textbooks such as Gazdar and Mellish (1989) and Allen (1995) provide an overview of this approach.

From the early nineties onwards, empirical methods based on corpus-based statistics, have gradually been re-introduced in the field, and have started to dominate it by the turn of the century, as can be seen from the number of papers subscribing to this approach in computational linguistics journals and conference proceedings. There are many reasons for this. Firstly, computer processing and storage capabilities have advanced to such an extent that statistical pattern recognition methods have become feasible on the large amounts of text and speech data that are now available in electronic form. Secondly, there has been an increase of interest within NLP (prompted by application-oriented and competitive funding) for the development of methods that scale well and can be used in real applications without requiring a complete syntactic and semantic analysis of text. Finally, simple statistical methods have been enormously successful in speech technology, and have therefore been applied to NLP as well. See Brill and Mooney (1997) and Church and Mercer (1993) for overviews of this empirical 'revolution' in NLP. The maturity of the approach is borne out by the publication of a few recent textbooks (Charniak, 1993, Manning and Schütze, 1999).

Comparing these empirical methods to the knowledge-based approach, it is clear that they have a number of advantages. In general, statistical approaches have a greater *coverage* of syntactic constructions and vocabulary, they are more robust (graceful degradation), they are reusable for different languages and domains, and development times for making applications and systems are shorter. On the other hand, knowledge-based methods allow the incorporation of linguistic knowledge and sophistication, making them sometimes more precise. Three crucial problems for (statistical) empirical methods are (i) the *sparse data problem*: often not enough data is available to estimate the probability of (low-frequency) events accurately, (ii) the *relevance problem*: it is often difficult to estimate reliably the importance or relevance of particular statistical events for the solution of the NLP problem, and (iii) the interpretation problem: most statistical techniques do not provide insight into *how* a trained statistical system solves a task.

The last few years have witnessed an increase of the use of symbolic machine learning methods in NLP. Some of these methods were created from within NLP (e.g. transformation-based error driven learning, Brill, 1995), other techniques were imported from Machine Learning into NLP; e.g. induction of decision trees and rules (Quinlan, 1993; Cohen, 1995), inductive logic programming (Lavrac and Dzeroski, 1994), and memory-based learning (Aha, 1997), Support Vector Machines (Vapnik, 1995). Recent collections of papers on Machine Learning of Natural Language are Wermter, Riloff, and Scheler (1996), Brill and Mooney (1997), and Cardie and Mooney (1999). These machine learning methods hold promise for solving the problems with statistical methods noted earlier. They incorporate new methods for smoothing data to solve sparse data problems and for assigning relevance to linguistic data, they allow the incorporation of linguistic background knowledge, and what they have learned is to a certain extent interpretable.

This paper introduces a set of studies describing *memory-based* approaches to NLP, one of these more recent additions to the suite of empirical techniques available to computational linguists, but with a rich history in other fields as we will see. Memory-Based Learning is inspired by the assumption that in learning a cognitive task from experience, people do not extract rules or other abstract representations from their experience, but reuse their memory of that experience directly. We will describe the theoretical background and inspiration sources of the approach, and attempt to put the different articles of this collection in perspective.

2 Inspiration Sources

Memory-Based learning and problem solving incorporates two principles: learning is the simple storage of experiences in memory, and solving a new problem is achieved by reusing solutions from *similar* previously solved problems.

For an example in the language processing field: in the well-known prepositional phrase disambiguation problem (PP-ATTACHMENT), where it has to be decided by a language understander which verb or noun is modified by a particular prepositional phrase, traces of usage of earlier similar cases may help in the disambiguation. E.g., in *eat a pizza with pineapple*, the prepositional phrase *with pineapple* modifies *pizza* rather than *eat* because we have memory traces of similar expressions (e.g. *eat pizza with anchovies*, *eat a sandwich with cheese*, ...) with the same noun-attachment. In *eat pizza with Eleni*, other memory traces of similar sentence fragments such as *eat crisps with Nicolas*, and *eat pizza with the boss* would favour a verb-modification interpretation. The feasibility of such an approach depends crucially on a good definition of *similarity* and the availability of sufficient examples.

This simple idea, and its many variants, has appeared regularly in work in Artificial

Intelligence, Psychology, Statistical Pattern Recognition, and Linguistics. This section describes the main inspiration sources for this type of algorithm. The next section situates the articles of this special issue within these traditions and in the context of related MBLP research.

2.1 Linguistics and Psycholinguistics

Since Chomsky replaced the vague notions of analogy and induction existing in linguistics in his time (in work of e.g. Saussure and Bloomfield) by the clearer and better operationalised notion of rule-based grammars, most mainstream linguistic theories, even the functionally and cognitively inspired ones, have assumed rules to be the only or main means to describe any aspect of language.

In contrast, the American linguist Royal Skousen (1989, 1992) argued for a specific operationalisation of the pre-Chomskyan analogical approach to language and language learning (AML, Analogical Modeling of Language). He introduced a definition of analogy that is not based on rules and that treats all language data at the same level without making a distinction between regular instances (obeying the rules) and irregular instances (exceptions to the rules). To model language acquisition and processing, a database of examples of language use is searched looking for instances analogous to a new item, and extrapolating a decision for the new item from them. The linguistic motivation for this approach is (i) the fact that in actual language use there is not a clear-cut all-or-none distinction between regular and irregular cases, (ii) the simplicity of the analogical approach as opposed to rule discovery, and (iii) the adaptability of the approach as opposed to the static, rigid rule-based alternative. Remarkably, seen from the outside, such an analogical approach appears to be rule-governed, and therefore adequately explains linguistic intuitions as well.

The specific analogical algorithm employed by Skousen is available in a number of implementations[1]. Current research attempts to solve the computational complexity problem (the algorithm is exponential in the number of attributes used to describe examples), and to apply the approach to a wide range of linguistic problems. The work has also been taken up as a psycholinguistically relevant explanation of human language acquisition and processing, especially as an alternative to *dual route* models of language processing (Eddington, 1998; Chandler, 1993; Derwing and Skousen, 1989). AML has also been used in computational linguistics. Jones (1996) describes an application of AML in Machine Translation, and Daelemans, Gillis, and Durieux (1997) compare AML to instance-based learning on a problem in computational phonology.

While AML is the most salient example of analogy-based theories in linguistics (and the most interesting from a computational linguistics point of view), other linguists outside the mainstream have proposed similar ideas. E.g. in the storage versus computation trade-off in models of linguistic processing, linguists like Bybee (1988), and linguistic theories such as Cognitive Grammar (Langacker, 1991) claim an important role for examples (instances of language use); but they still presuppose rules to be essential for representing generalizations.

It is interesting to see that also in general psychology, in studies of human categorization, exemplar-based models often produce good fits of human behaviour and errors (Smith and Medin, 1981, Nosofsky, 1986). These models assume that people represent categories by storing individual exemplars in memory, and make categorization decisions based on the similarity of stimuli to these stored exemplars. They are contrasted with prototype-based, probabilistic or classical 'rule-based' categorization models.

[1] See the AML group's homepage at `http://humanities.byu.edu/aml/homepage.html`.

2.2 Statistical Pattern Recognition and Artificial Intelligence

As far as the algorithms used in MBLP are concerned, nearest neighbor methods (k-nn), developed in statistical pattern recognition from the fifties onwards, have played an important inspirational role (e.g. Fix and Hodges, 1951; Cover and Hart, 1967). In these methods, examples (labeled with their class) are represented as points in an example space with as dimensions the numeric attributes used to describe the examples. A new example obtains its class by finding its position as a point in this space, and extrapolating its class from the k nearest points in its neighbourhood. Nearness is defined as the reverse of Euclidean distance. A very early citation nicely capturing the intuitive attraction of the k-nn approach is the following:

> "This "rule of nearest neighbor" has considerable elementary intuitive appeal and probably corresponds to practice in many situations. For example, it is possible that much medical diagnosis is influenced by the doctor's recollection of the subsequent history of an earlier patient whose symptoms resemble in some way those of the current patient." (Fix and Hodges, 1952)

This literature has also generated many studies on methods for removing examples from memory either for efficiency (faster processing by removing unnecessary examples) or for accuracy (better predictions for unseen cases by removing badly predicting examples). See Dasarathy (1991) for a collection of fundamental papers on k-nn research.

However, until recently, the impact of these non-parametric statistical methods on the development of systems for solving practical problems has remained limited because of a number of shortcomings: they were computationally expensive in storage and processing; intolerant of attribute noise and irrelevant attributes; sensitive to the similarity metric used; and the Euclidean distance metaphor for similarity breaks down with non-numeric and missing feature values.

From the late eighties onwards, the intuitive appeal of the nearest neighbor approach has been adopted in Artificial Intelligence in many variations on the basic nearest neighbor modeling idea, using names such as memory-based reasoning, case-based reasoning, exemplar-based learning, locally-weighted learning, and instance-based learning (Stanfill and Waltz, 1986; Cost and Salzberg, 1993; Riesbeck and Schank, 1989; Kolodner 1993; Atkeson, Moore, and Schaal, 1997; Aamodt and Plaza, 1994; Aha, Kibler, and Albert, 1991). These methods modify or extend the nearest neighbor algorithm in different ways, and aim to solve (some of) the problems with k-nn listed before.

Recently, the term *Lazy Learning* (as opposed to *eager learning*) has been proposed for this family of methods (Aha, 1997) because all these methods (i) defer processing of input until needed, (ii) process input by combining stored data, and (iii) discard processed input afterwards. These methods often yield highly adaptive behavior, and there have been many successful applications of these approaches in robotics, control, vision, problem solving, reasoning, decision making, diagnosis, information retrieval, and data mining (see e.g. Kasif et al., 1997).

3 Memory-Based Language Processing Literature

Given the long tradition of analogical approaches in linguistics (even if not in the mainstream), their potential psychological relevance, and the success of memory-based methods in pattern recognition and AI applications, it is not surprising that the approach has also surfaced in Natural Language Processing. Apart from the advantages inherent in all learning approaches, as discussed earlier (fast development, robustness, high coverage, etc.), advantages commonly associated with a memory-based approach to NLP include ease of learning (simply storing examples), ease of integrating multiple sources of information, and the use of similarity-based reasoning as a smoothing method for estimating low-frequency events. Especially the last property is an important theoretical issue. In language processing tasks, unlike other typical AI tasks, low-frequency events are pervasive. Due to borrowing, historical change, and the complexity of language, most data sets representing NLP tasks contain few regularities, and many subregularities and exceptions. It is impossible for inductive algorithms to reliably distinguish noise from exceptions, so non-abstracting lazy memory-based learning algorithms should be at an advantage compared to eager learning methods such as decision tree learning or rule induction: 'forgetting exceptions is harmful'. The usefulness of similarity for smoothing is discussed in Zavrel and Daelemans (1997) and Dagan, Lee and Pereira (1999). Daelemans, van den Bosch, and Zavrel (1999) provide empirical results and theoretical analysis supporting the 'forgetting exceptions is harmful' hypothesis. **Antal van den Bosch (this volume)** takes this analysis further by studying different methods of bottom-up abstraction from instances on a large range of NLP problems, and shows that it still holds in general. However, limited, careful abstraction using a notion of *instance families* (implemented in the FAMBL algorithm) can be used to prune memory without adverse effects on generalization accuracy on small data sets.

3.1 Memory-Based Computational Linguistics

Since the early nineties, we find several studies using nearest-neighbour techniques for solving NLP disambiguation problems, framed as classification problems. These tasks define modules that can play a role in different concrete applications. Each module is trained by collecting a set of examples with the required features for that module. E.g., morphosyntactic disambiguation (part-of-speech tagging) can be solved by representing information about the form of the word to be disambiguated and about the words in its immediate context as features, and the correct morphosyntactic category of the word in that context as class. New words in context are assigned a class on the basis of the nearest neighbors. Such a tagger can then be used as a component of applications such as information extraction or translation, or several of such components can be combined to perform a more complex task, such as a parser.

Cardie (1993a, 1994) addresses case-based lexical, semantic, and structural disambiguation of full sentences in limited domains, co-reference and anaphora resolution (Cardie, 1996).

Daelemans and colleagues in Antwerp and Tilburg have applied a specific approach to MBLP (based on global feature weighting, IB1-IG, and tree indexing for efficiency, IGTREE) to a large number of NLP tasks: hyphenation and syllabification (Daelemans and van den Bosch, 1992); assignment of word stress (Daelemans, Durieux, and Gillis, 1994); grapheme-to-phoneme conversion (Daelemans and Van den Bosch, 1996); morphological analysis (Van den Bosch and Daelemans, 1999); part-of-speech tagging (Daelemans, Zavrel, Berck, and Gillis, 1996); prepositional phrase attachment (Zavrel, Daelemans, and Veenstra, 1997); word sense disambiguation (Veenstra et al., 1999); shallow parsing (Buchholz, Veenstra, and Daelemans, 1999). A partial overview paper is (Daelemans, 1995). The algorithms used are described and reviewed in Daelemans, Van den Bosch, and Weijters (1997), and in the documentation of the freely available TiMBL package implementing a large range of memory-based algorithms[2].

Lehnert (1987), and Weijters (1991) are early examples of memory-based learning applied to grapheme-to-phoneme conversion. Ng and Lee (1996), and Fujii, Inui, Tokunaga, and Tanaka (1998) also apply memory-based techniques to the problem of Word Sense Disambiguation. Similar nearest-neighbour-inspired approaches have been applied

[2]Available from `http://ilk.kub.nl`. Papers in electronic form and demonstrations are available from that web-site as well.

to context-sensitive parsing (Simmons and Yu, 1993), and machine translation (Hermjakob, 1997; Hermjakob and Mooney, 1997). There are also memory-based approaches to text categorization and filtering (Masand, Linoff, and Waltz, 1992; Yang and Chute, 1994; Riloff and Lehnert, 1994).

One especially crucial problem for these approaches is the weighting of the relevance of the features in solving the task. Giving equal weight to all features (as in the basic k-nn algorithm) while computing the similarity of a new case to examples in memory, would overrate the importance of irrelevant and redundant features. Cardie (1993b) uses the information gain splitting criterion used in decision tree learning (Quinlan, 1993) to *select* relevant features, whereas Daelemans and Van den Bosch (1992) use the same information-theoretic technique to globally *weight* feature relevance (the IB1-IG algorithm). Feature weighting is currently a topic of intensive investigation in lazy learning research. Wettschereck, Aha, and Mohri (1997) provide a recent review. **Cardie (this volume)** shows that psychological constraints such as recency effects and short term memory limitations can be integrated in case-based learning for successful feature selection and weighting.

3.2 Data-Oriented Parsing

DOP (Data-Oriented Parsing) is a memory-based approach to syntactic parsing (Scha, 1992; Bod, 1995, 1998; Bod and Scha, 1997; Bonnema, Bod, and Scha, 1997) which uses a corpus of parsed or semantically analyzed utterances (a treebank) as a representation of a person's language experience, and analyzes new sentences searching for a recombination of subtrees that can be extracted from this treebank. The frequencies of these subtrees in the corpus are used to compute the probability of analyses. Such a method uses an annotated corpus as grammar, an approach formalized as Stochastic Tree Substitution Grammar (STSG). The advantage of STSG is that lexical information and idiomatic expressions (multi-word lexical items) can in principle play a role in finding and ranking an analysis. **Scha, Bod, and Sima'an (this volume)** provide an in-depth overview of the approach, tracing its motivation in pre-Chomskyan linguistics, its computational and optimization aspects, experimental results, and a thorough comparison with other memory-based approaches.

Recently, a new memory-based sentence analysis method, Memory-Based Sequence Learning (MBSL) was introduced by Argamon, Dagan, and Krymolowski (1998) that is reminiscent of both DOP and the nearest neighbor approach. It shares with DOP the ability to take into account all substrings of an analyzed string and their frequency in similarity-based extrapolation, and it shares with nearest-neighbor classification-based approaches the modular set-up (one MBSL system for each task). **Argamon, Dagan, and Krymolowski (this volume)** contains a thorough discussion of this algorithm and empirical results on shallow parsing.

3.3 Example-Based Machine Translation (EBMT)

In seminal work, Nagao (1984) proposed an approach to Machine Translation which is essentially memory-based. By storing a large set of (analyzed) sentences or sentence fragments in the source language with their associated translation in the target language as examples, a new source language sentence can be translated by finding examples in memory that are *similar* to it in terms of syntactic structure and word meaning, and extrapolating from the translations associated with these examples. Jones (1996) provides an overview of different approaches within EBMT since Nagao (1984). In practice, because of the huge space of possible sentences to be translated, and the cost of collecting and searching large amounts of examples, EBMT systems are mostly hybrid, and contain rule-based as well as memory-based components. **Andy Way (this volume)** proposes and analyzes a specific instance of such a hybrid approach. He critically evaluates LFG-based machine translation, and data-oriented translation, and shows that a combination of both can be extended to serve as a novel hybrid model for Machine Translation. It is interesting to note that *translation memories*, arguably the most successful commercial approach to machine-aided translation today, are also based on the memory-based framework: large amounts of documents aligned with their translations are stored, and a possible translation for a new sentence is searched using a fuzzy string matching technique.

3.4 Analogy and Similarity

The concept of analogy implicit in memory-based classification approaches based on a nearest-neighbor metaphor is simple and empirically adequate for many language processing tasks, but it has been argued that a linguistically adequate approach requires a more sophisticated concept of analogy, especially to handle the non-compositional, holistic aspects of language. *Paradigm-based proportional analogy* is such an algorithmic definition of the classical linguistic notion of proportional analogy (Pirrelli and Federici, 1993, 1994; Yvon, 1997; Lepage, 1998). **Pirrelli and Yvon (this volume)**, provide a synthesis of this approach to analogy-based natural language learning, and show its merits on a large range of language processing tasks.

4 Conclusion

In this short paper, I have tried to sketch how two traditions: analogy-based language models in linguistics, and nearest-neighbour-based learning methods in AI, have merged in what we have called *memory-based language processing*. Current incarnations of this approach range from nearest-neighbor classification for language processing tasks, over memory-based structure learning and parsing, to example-based machine translation, and linguistically motivated computational models of analogical language learning. All these approaches share the same underlying idea of solving problems from experience directly rather than from some knowledge structure extracted from experience. To conclude this introduction, table 1 provides an overview of the themes and tasks addressed by the different authors publishing in this special issue.

5 Acknowledgements

The origin of this special issue is a workshop on memory-based approaches to language processing held December 1997 at Corsendonk abbey, Turnhout, Belgium (yes, the same 'abbey' where the infamous *Monks problems* originate from). The feeling of the participants was that a collection of papers on MBLP, based on an open call for papers, would be useful and timely. Many thanks to Steven Gillis (with whom I organized the workshop), and to Royal Skousen, Dave Waltz, David Aha, Ton Weijters, Remko Scha, Antal van den Bosch, and the other participants of the workshop for their support and encouragement for this idea. Thanks to Eric Dietrich for providing a medium for the realisation of this collection. Many thanks also to the 17 reviewers of this special issue for their help, and to the authors for their cooperation and their enjoyable submissions. My work on this special issue has been made possible partially by grants from FWO (Belgium) and NWO (The Netherlands).

Cardie	feature selection	relative pronoun disambiguation semantic class prediction part of speech prediction
Van den Bosch	weak abstraction	noun phrase recognition prepositional phrase attachment part of speech tagging word pronunciation stress assignment morphological analysis
Argamon, Dagan Krymolowski	structure learning	noun phrase recognition subject-verb relations verb-object relations
Scha, Bod, Sima'an	DOP	full parsing
Way	LFG-DOP	machine translation
Pirrelli, Yvon	paradigm-based proportional analogy	word pronunciation morphological analysis word sense disambiguation

Table 1: Themes and tasks addressed by the articles in this special issue.

6 References

Aamodt, A. and Plaza, E., 1994, Case-Based Reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications* 7, 39–59.

Aha, D. (ed.), 1997, *Lazy Learning* (Dordrecht: Kluwer Academic Publishers).

Aha, D., Kibler, D., and Albert, M., 1991, Instance-based learning algorithms. *Machine Learning*, 7, 37–66.

Allen, J., 1995, *Natural Language Understanding* (Redwood City: Benjamin/Cummings).

Argamon, S., Dagan, I., and Krymolowski, Y., 1998, A memory-based approach to learning shallow natural language patterns, In *Proc. of COLING/ACL*, pp. 67–73, Montreal, Canada.

Atkeson, C., Moore, A., and Schaal, S., 1997, Locally Weighted Learning. *Artificial Intelligence Review* 11: 11–73.

Bod, R., 1995, *Enriching Linguistics with Statistics: Performance Models of Natural Language*, Ph.D. thesis, ILLC Dissertation Series 1995-14, University of Amsterdam.

Bod, R., 1998, *Beyond Grammar* (Cambridge: CSLI / Cambridge University Press).

Bod, R., and Scha, R., 1997, Data-oriented language processing. In Young, S. and Bloothooft, G. (eds.) *Corpus-Based Methods in Language and Speech Processing* (Boston: Kluwer Academic Publishers), pp. 137–173.

Bonnema, R., Bod, R., and Scha, R., 1997, A DOP Model for Semantic Interpretation, In *Proc. of the 35th annual meeting of the ACL*, pp. 159–167, Madrid, Spain.

Bosch, A. van den, and Daelemans, W., 1999, Memory-Based Morphological Analysis, In *Proc. of the 37th annual meeting of the ACL*, pp. 285–292, Maryland, USA.

Brill, E., 1995, Transformation-based Error-driven Learning and Natural Language Processing: A Case Study in Part-of-Speech tagging. *Computational Linguistics*, 21(4): 543–565.

Brill, E., and Mooney, R., 1997, An overview of empirical natural language processing. *AI Magazine* 18(4): 13–24.

Brill, E., and Mooney, R. (guest eds.), 1997, Special Issue on Empirical Natural Language Processing. *AI Magazine*, 18(4).

Buchholz, S., Veenstra, J., and Daelemans, W., 1999, Cascaded Grammatical relation Assignment. *Proc. of the joint SIGDAT Conference on Empirical Methods in Latural Language Processing and Very Large Corpora.*, pp. 239–246, Maryland, USA.

Bybee, J., 1988, Morphology as Lexical organization. In M. Hammond and M. Noonan (eds.), *Theoretical morphology. Approaches in modern linguistics* pp. 119–141, (San Diego: Academic Press).

Cardie, C., 1993a, A case-based approach to knowledge acquisition for domain-specific sentence analysis. In *Proc. of the 11th national conference on Artificial Intelligence*, 798–803, Washington, DC.

Cardie, C., 1993b, Using decision trees to improve case-based learning. In P. Utgoff (ed.) *Proc. of the tenth international conference on machine learning*, pp. 25–32, Amherst, MA.

Cardie, C., 1994, *Domain-specific knowledge acquisition for conceptual sentence analysis.* Ph.D. thesis, University of Massachusetts, Amherst, MA.

Cardie, C., 1996, Automatic feature set selection for case-based learning of linguistic knowledge. *Proc. of the 1st Conference on Empirical Methods in Natural Language Processing*, pp. 113–126.

Cardie, C., and Mooney, R. (guest eds.), 1999, Machine Learning and Natural Language. Special Issue of *Machine Learning*, 11:(1-3).

Chandler, S., 1993, Are rules and modules really necessary for explaining language? *Journal of Pyscholinguistic Research* 22: 593–606.

Charniak, E., 1993, *Statistical Language Learning* (Cambridge, MA: The MIT Press).

Church, K., and Mercer, R., 1993, Introduction to the Special Issue on Computational Linguistics Using Large Corpora. *Computational Linguistics*, 19(1): 1–24.

Cost, S. and Salzberg, S., 1993, A weighted nearest neighbor algorithm for learning with symbolic features. *Machine Learning* 10(1): 57–78.

Cohen, W., 1995, Fast effective rule induction. In *Proc. of the Twelfth International Conference on Machine Learning*, Lake Tahoe, California.

Cover, T., and Hart, P., 1967, Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13: 21–27.

Daelemans, W., and van den Bosch, A., 1992, Generalisation performance of backpropagation learning on a syllabification task. In M. Drossaers and A. Nijholt (eds.) *TWLT3: Connectionism and Natural Language Processing* (Enschede: Twente University), pp. 27–37.

Daelemans, W., Durieux, G., and Gillis, S., 1994, The acquisition of stress: A data-oriented approach. *Computational Linguistics*, 20(3): 421–451.

Daelemans, W., 1995, Memory-based lexical acquisition and processing. In P. Steffens (ed) *Machine Translation and the Lexicon*, volume 898 of *Lecture Notes in Artificial Intelligence* (Berlin: Springer-Verlag), pp. 85–98.

Daelemans, W. and Van den Bosch, A. Language-independent data-oriented grapheme-to-phoneme conversion. In J. P. H. Van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg (eds), *Progress in Speech Synthesis*, (Berlin: Springer-Verlag), pp. 77–89.

Daelemans, W., Zavrel, J., Berck, P., Gillis, S., 1996, MBT: A memory-based part of speech tagger generator. In E. Ejerhed and I. Dagan (eds) *Proc. of the Fourth Workshop on Very Large Corpora*, pp. 14–27.

Daelemans, W., Gillis, S., and Durieux, G., 1997, Skousen's Analogical Modeling Algorithm: A comparison with Lazy Learning. In D. Jones and H. Somers (eds.) *New Methods in Language Processing* (London: UCL Press), pp. 3–15.

Daelemans, W., Van den Bosch, A., and Weijters, A., 1997, IGTREE: using trees for compression and classification in lazy learning algorithms. *Artificial Intelligence Review*, 11: 407–423.

Daelemans, W., van den Bosch, A., and Zavrel, J., 1999, Forgetting exceptions is harmful in language learning. *Machine Learning* 34: 11–41.

Dagan, I., Lee, L., and Pereira, F., 1999, Similarity-Based Models of Word Cooccurrence. *Machine Learning* 34: 43–69.

Dasarathy, B. (ed.), 1991, *Nearest neighbor (NN) norms: NN pattern classification techniques* (Los Alamitos, CA: IEEE Computer Society Press).

Derwing, B., and Skousen, R., 1989, Morphology in the Mental Lexicon: A New Look at Analogy. *Morphology Yearbook* 2: 55–71.

Eddington, D., 1998, Applying Skousen's analogical approach to Spanish nominals in -ion. In R. Smyth (ed) *Festschrift for Bruce L. Derwing*, to appear.

Fix, E., and Hodges, J., 1951, Discriminatory Analysis: Nonparametric Discrimination: Consistency properties. Project 21-49-004, Report number 4, pp. 261–279, USAF School of Aviation Medicine, Randolph Field, Texas.

Fix, E., and Hodges, J., 1952, Discriminatory Analysis: Nonparametric Discrimination: Small Sample Performance. Project 21-49-004, Report number 11, pp. 280–322, USAF School of Aviation Medicine, Randolph Field, Texas.

Fujii, A., Inui, K., Tokunaga, T., and Tanaka, H., 1998, Selective sampling for example-based word sense disambiguation. *Computational Linguistics* 24(4): 573–598.

Gazdar, G., and Mellish, C., 1989, *Natural Language Processing in LISP* (Reading, MA: Addison-Wesley).

Hermjakob, U., 1997, *Learning parse and translation decisions from examples with rich context.* Ph.D. dissertation, University of Texas at Austin, AI97-261.

Hermjakob, U., and Mooney, R., 1997, Learning parse and translation decisions from examples with rich context. In *Proc. of the 35th Annual Meeting of the ACL*, pp. 482–489, Madrid, Spain.

Jones, D., 1996, *Analogical Natural Language Processing*, (London: UCL Press).

Kasif, S., Salzberg, S., Waltz, D., Rachlin, J., Aha, D., 1997, Towards a framework for Memory-Based Reasoning. Submitted.

Kolodner, J., 1993, *Case-Based Reasoning* (San Mateo: Morgan Kaufmann).

Langacker, R., 1991, *Concept, Image, and Symbol. The Cognitive Basis of Grammar* (Berlin: Mouton De Gruyter).

Lavrac, N., and Dzeroski, S., 1994, *Inductive Logic Programming: Techniques and Applications* (Chichester: Ellis Horwood).

Lehnert, W., 1987, Case-based problem solving with a large knowledge base of learned cases. In *Proc. of the sixth national conference on Artificial Intelligence*, pp. 301–306, Los Altos, CA.

Lepage, Y., 1998, Solving analogies on words: an algorithm. In *Proc. of the 36th annual meeting of the ACL*, pp. 718–735, Montreal, Canada.

Manning, C., and Schütze, H., 1999, *Foundations of Statistical Natural Language Processing* (Cambridge, MA: The MIT Press).

Masand, B., Linoff, G., and Waltz, D., 1992, Classifying news stories using memory-based reasoning. In *Proc. SIGIR*, pp. 59–65.

Nagao, M., 1984, A framework of a mechanical translation between Japanese and English by analogy principle. In A. Elithorn and R. Banerji (eds.) *Artificial and human intelligence* (Amsterdam: North-Holland), pp. 173–180.

Ng, H. and Lee, H., 1996, Integrating multiple knowledge sources to disambiguate word sense: an exemplar-based approach. In *Proc. of 34th Annual Meeting if the Association for Computational Linguistics*, pp. 40–47.

Nosofsky, R., 1986, Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 15: 39–57.

Pirelli, V., and Federici, S., 1993, An analogical way to language modelling: Morpheus. *Acta Linguistica Hungarica*, 41: 235–263.

Pirrelli, V., and Federici, S., 1994, "Derivational" paradigms in morphonology. In *Proc. of the 16th International Conference on Computational Linguistics*, pp. 234–240, Kyoto, Japan.

Quinlan, J.R., 1993, C4.5*: Programs for Machine Learning* (San Mateo, CA: Morgan Kaufmann).

Riesbeck, C., and Schank, R., 1989, *Inside Case-Based Reasoning* (Northvale, NJ: Erlbaum).

Riloff, E., and Lehnert, W., 1994, Information Extraction as a basis for high-precision text classification. *ACM Transactions on Information Systems*, 12(3): 296–333.

Scha, R., 1992, Virtuele Grammatica's en Creatieve Algoritmen. *Gramma/TTT* 1(1): 57–77.

Simmons, R., and Y. Yu, 1993, The acquisition and use of context-dependent grammars for English. *Computational Linguistics*, 18(4): 391–418.

Skousen, R., 1989, *Analogical Modeling of Language* (Dordrecht: Kluwer).

Skousen, R., 1992, *Analogy and Structure* (Dordrecht: Kluwer).

Smith, E. and Medin, D., 1981, *Categories and Concepts* (Cambridge, MA: Harvard University Press).

Stanfill, C., and Waltz, D., 1986, Toward memory-based reasoning. *Communications of the ACM*, 29: 1212–1228.

Vapnik, V., 1995, *Estimation of dependencies based on empirical data.* Springer Series in Statistics (Berlin: Springer).

Veenstra, J., Van den Bosch, A., Buchholz, S., Daelemans, W., and Zavrel, J., 1999, Memory-Based Word Sense Disambiguation, submitted.

Weijters, A., 1991, A simple look-up procedure superior to NETtalk? *Proc. of the international conference on artificial neural networks.*, Espoo, Finland.

Wermter, S., Riloff, E., and Scheler, G. (eds.), 1996, *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, Springer Lecture Notes in Artificial Intelligence 1040 (Berlin: Springer).

Wettschereck, D., Aha, D., and Mohri, T., 1997, A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review* 11: 273–314.

Yang, Y., and Chute, C., 1994, An example-based mapping method for text classification and retrieval. *ACM Transactions on Information Systems*, 12(3): 252–295.

Yvon, F., 1997, Paradigmatic cascades: a linguistically sound model of pronunciation by analogy. In '*Proc. of the 35th annual meeting of the ACL*, pp. 428–435, Madrid, Spain.

Zavrel, J. and Daelemans, W., 1997, Memory-based learning: Using similarity for smoothing. *Proc. of the 35th annual meeting of the ACL*, pp. 436–442, Madrid, Spain.

Zavrel, J., Daelemans, W., and Veenstra, J., 1997, Resolving PP attachment ambiguities with memory-based learning. In *Proc. of the workshop on computational natural language learning*, pp. 136–144, Madrid, Spain.