

Automated Induction of Sense in Context

James PUSTEJOVSKY, Patrick HANKS, Anna RUMSHISKY

Brandeis University

Waltham, MA 02454, USA

{jamesp,patrick,arum}@cs.brandeis.edu

Abstract

In this paper, we introduce a model for sense assignment which relies on assigning senses to the contexts within which words appear, rather than to the words themselves. We argue that word senses as such are not directly encoded in the lexicon of the language. Rather, each word is associated with one or more stereotypical syntagmatic patterns, which we call *selection contexts*. Each selection context is associated with a meaning, which can be expressed in any of various formal or computational manifestations. We present a formalism for encoding contexts that help to determine the semantic contribution of a word in an utterance. Further, we develop a methodology through which such stereotypical contexts for words and phrases can be identified from very large corpora, and subsequently structured in a selection context dictionary, encoding both stereotypical syntactic and semantic information. We present some preliminary results.

1 Introduction

This paper describes a new model for the acquisition and exploitation of selectional preferences for predicates from natural language corpora. Our goal is to apply this model in order to construct a dictionary of normal *selection contexts* for natural language; that is, a computational lexical database of rich selectional contexts, associated with procedures for assigning interpretations on a probabilistic basis to less normal contexts. Such a semi-automatically developed resource promises to have applications for a number of NLP tasks, including word-sense disambiguation, selectional preference acquisition, as well as anaphora resolution and inference in specialized domains. We apply this methodology to a selected set of verbs, including a subset of the verbs in the Senseval 3 word sense discrimination task and report our initial results.

1.1 Selectional Preference Acquisition: Current State of the Art

Predicate subcategorization information constitutes an essential part of the computational lexicon entry. In recent years, a number of approaches have been proposed for dealing computationally with selectional preference acquisition (Resnik (1996); Briscoe and Carroll (1997); McCarthy (1997); Rooth et al.

(1999); Abney and Light (1999); Ciaramita and Johnson (2000); Korhonen (2002)).

The currently available best algorithms developed for the acquisition of selectional preferences for predicates are induction algorithms modeling selectional behavior as a distribution over words (cf. Abney and Light (1999)). Semantic classes assigned to predicate arguments in subcategorization frames are either derived automatically through statistical clustering techniques (Rooth et al. (1999), Light and Greiff (2002)) or assigned using hand-constructed lexical taxonomies such as the WordNet hierarchy or LDOCE semantic classes. Overwhelmingly, WordNet is chosen as the default resource for dealing with the sparse data problem (Resnik (1996); Abney and Light (1999); Ciaramita and Johnson (2000); Agirre and Martinez (2001); Clark and Weir (2001); Carroll and McCarthy (2000); Korhonen and Preiss (2003)).

Much of the work on inducing selectional preferences for verbs from corpora deals with predicates indiscriminately, assuming no differentiation between predicate senses (Resnik (1996); Abney and Light (1999); Ciaramita and Johnson (2000); Rooth et al. (1999)). Those approaches that do distinguish between predicate senses or complementation patterns in acquisition of selectional constraints (Korhonen (2002); Korhonen and Preiss (2003)) do not use corpus analysis for verb sense classification.

1.2 Word Sense Disambiguation: Current State of the Art

Previous computational concerns for economy of grammatical representation have given way to models of language that not only exploit generative grammatical resources but also have access to large lists of contexts of linguistic items (words), to which new structures can be compared in new usages. However, following the work of Yarowsky (1992), Yarowsky (1995), many supervised WSD systems use minimal information about syntactic structures, for the most part restricting the notion of context to topical and local features. Topical features track open-class words that appear within a certain window around a target word, and local features track small N-grams associated with the target word. Disambiguation therefore relies on word co-occurrence statistics, rather than on structural similarities. That remains the case for most systems

that participated in Senseval-2 (Preiss and Yarowsky (2001)). Some recent work (Stetina et al. (1998); Agirre et al. (2002); Yamashita et al. (2003)) attempts to change this situation and presents a directed effort to investigate the impact of using syntactic features for WSD learning algorithms. Agirre et al (2002) and Yamashita et al. (2003) report resulting improvement in precision.

Stevenson and Wilks (2001) propose a somewhat related technique to handle WSD, based on integrating LDOCE classes with simulated annealing. Although space does not permit discussion here, initial comparisons suggest that our selection contexts could incorporate similar knowledge resources; it is not clear what role model bias plays in associating patterns with senses, however.

In this paper we modify the notion of word sense, and at the same time revise the manner in which senses are encoded. The notion of word sense that has been generally adopted in the literature is an artifact of several factors in the status quo, notably the availability of lexical resources such as machine-readable dictionaries, in which fine sense distinctions are not supported by criteria for selecting one sense rather than another, and WordNet, where *synset* groupings are taken as defining word sense distinctions. Thus, for instance, Senseval-2 WSD tasks required disambiguation using WordNet senses (see, e.g., discussion in Palmer et al. (2004)). The feature sets used in the supervised WSD algorithms at best use only minimal information about the typing of arguments. The approach we adopt, *Corpus Pattern Analysis (CPA)* (Pustejovsky and Hanks (2001)), incorporates semantic features of the arguments of the target word. Semantic features are expressed in terms of a restricted set of shallow types, chosen for their prevalence in selection context patterns. This type system is extended with predicate-based noun clustering, in the bootstrapping process described below.

1.3 Related Resources: FrameNet

It is necessary to say a few words about the differences between CPA and FrameNet. The CPA approach has its origins in the analysis of large corpora for lexicographic purposes (e.g. Cobuild (Sinclair et al., 1987)) and in systemic-functional grammar, in particular in Halliday’s notion of “lexis as a linguistic level” (Halliday, 1966) and Sinclair’s empirical approach to collocational analysis (Sinclair, 1991). FrameNet (freely available online in a beautifully designed data base at <http://www.icsi.berkeley.edu/~framenet/>), is an attempt to implement Fillmore’s 1975 proposal that, instead of seeking to satisfy a set of necessary and sufficient conditions, the meanings of words in text should be analyzed by calculating resemblance to a prototype (Fillmore, 1975).

CPA (Hanks, 2004) is concerned with establishing prototypical norms of usage for individual words. It is possible (and certainly desirable) that CPA norms

will be mappable onto FrameNet’s semantic frames (for which see the whole issue of the International Journal of Lexicography for September 2003 (in particular Atkins et al. (2003a), Atkins et al. (2003b), Fillmore et al. (2003a), Baker et al. (2003), Fillmore et al. (2003b)). In frame semantics, the relationship between semantics and syntactic realization is often at a comparatively deep level, i.e. in many sentences there are elements that are potentially present but not actually expressed. For example, in the sentence “he risked his life”, two semantic roles are expressed (the risker and the valued object “his life” that is put at risk). But at least three other roles are subliminally present although not expressed: the possible bad outcome (“he risked his death”), the beneficiary or goal (“he risked his life for her/for a few dollars”), and the means (“he risked a backward glance”).

CPA, on the other hand, is shallower and more practical: the objective is to identify, in relation to a given target word, the overt textual clues that activate one or more components of its meaning potential. There is also a methodological difference: whereas FrameNet research proceeds frame by frame, CPA proceeds word by word. This means that when a word has been analysed in CPA the patterns are immediately available for disambiguation. FrameNet will be usable for disambiguation only when all frames have been completely analysed. Even then, FrameNet’s methodology, which requires the researchers to think up all possible members of a Frame a priori, means that important senses of words that have been partly analysed are missing and may continue to be missing for years to come. There is no attempt in FrameNet to identify the senses of each word systematically and contrastively. In its present form, at least, FrameNet has at least as many gaps as senses. For example, at the time of writing *toast* is shown as part of the *Apply_Heat* frame but not the *Celebrate* frame. It is not clear how or whether the gaps are to be filled systematically. We do not even know whether there is (or is going to be) a *Celebrate* frame and if so what it will be called. What is needed is a principled fix – a decision to proceed from evidence, not frames. This is ruled out by FrameNet for principled reasons: the unit of analysis for FrameNet is the frame, not the word.

2 CPA Methodology

The Corpus Pattern Analysis (CPA) technique uses a semi-automatic bootstrapping process to produce a dictionary of selection contexts for predicates in a language. Word senses for verbs are distinguished through corpus-derived syntagmatic patterns mapped to Generative Lexicon Theory (Pustejovsky (1995)) as a linguistic model of interpretation, which guides and constrains the induction of senses from word distributional information. Each pattern is specified in terms of lexical sets for each argument, shallow semantic typing of these sets, and

other syntagmatically relevant criteria (e.g., adverbials of manner, phrasal particles, genitives, negatives).

The procedure consists of three subtasks: (1) the manual discovery of selection context patterns for specific verbs; (2) the automatic recognition of instances of the identified patterns; and (3) automatic acquisition of patterns for unanalyzed cases. Initially, a number of patterns are manually formulated by a lexicographer through corpus pattern analysis of about 500 occurrences of each verb lemma. Next, for higher frequency verbs, the remaining corpus occurrences are scrutinized to see if any low-frequency patterns have been missed. The patterns are then translated into a feature matrix used for identifying the sense of unseen instances for a particular verb. In the remainder of this section, we describe these subtasks in more detail. The following sections explain the current status of the implementation of these tasks.

2.1 Lexical Discovery

Norms of usage are captured in what we call selection context patterns. For each lemma, contexts of usage are sorted into groups, and a stereotypical CPA pattern that captures the relevant semantic and syntactic features of the group is recorded. Many patterns have alternations, recorded in satellite CPA patterns. Alternations are linked to the main CPA pattern through the same sense-modifying mechanisms as those that allow for exploitations (coercions) of the norms of usage to be understood. For example, here is the set of patterns for the verb *treat*. Note that these patterns do not capture all possible uses, and other patterns may be added, e.g. if additional evidence is found in domain-specific corpora.

(1) CPA Pattern set for *treat*:

- I. `[[Person 1]] treat [[Person 2]] {{at | in} [[Location]] (for [[Event = Injury | Ailment]]); NO [Adv[Manner]]}`
- II. `[[Person 1]] treat [[Person 2]] [Adv[Manner]]`
- IIIa. `[[Person]] treat [[TopType 1]] {{as | like} [[TopType 2]]}`
- IIIb. `[[Person]] treat [[TopType]] {{as if | as though | like} [CLAUSE]}`
- IV. `[[Person 1]] treat [[Person 2]] {to [[Event]]}`
- V. `[[Person]] treat [[PhysObj | Stuff 1]] (with [[Stuff 2]])`

There may be several patterns realizing a single sense of a verb, as in (IIIa/IIIb) above. Also, there may be several equivalent alternations or there may be a stereotype. Note that alternations are different realizations of the same norm, not exploitations (i.e., not coercions).

(2) Alternations for *treat* Pattern 1 :

- `[[Person 1]] treat [[Person 2]] {{at | in} [[Hospital]] (for [[Injury | Ailment]]); NO [Adv[Manner]]}`
- Alternation 1:
`[[Person 1 <--> Medicament | Med-Procedure | Institution]]`
- Alternation 2:
`[[Person 2 <--> Injury | Ailment | Bodypart]]`

CPA Patterns

A CPA pattern extends the traditional notion of selectional context to include a number of other contextual features, such as minor category parsing and subphrasal cues. Accurate identification of the semantically relevant aspects of a pattern is not an obvious and straightforward procedure, as has sometimes been assumed in the literature. For example, the presence or absence of an adverbial of manner in the third valency slot around a verb can dramatically alter the verb's meaning. Simple syntactic encoding of argument structure, for instance, is insufficient to discriminate between the two major senses of the verb *treat*, as illustrated below.

- (3) a. They say their bosses **treat** them with respect.
- b. Such patients are **treated** with antibiotics.

The ability to recognize the shallow semantic type of a phrase in the context of a predicate is of course crucial—for example, in (3a) recognizing the PP as (a) an adverbial, and (b) an adverbial of manner, rather than an instrumental co-agent (as in (3b)), is crucial for assigning the correct sense to the verb *treat* above.

In the CPA model, automatic identification of selection contexts not only captures the argument structure of a predicate, but also more delicate features, which may have a profound effect on the semantic interpretation of a predicate in context. There are four constraint sets that contribute to the patterns for encoding selection contexts. These are:

- (4) a. *Shallow Syntactic Parsing*: Phrase-level recognition of major categories.
- b. *Shallow Semantic Typing*: 50-100 primitive shallow types, such as Person, Institution, Event, Abstract, Artifact, Location, and so forth. These are the top types selected from the Brandeis Shallow Ontology (BSO), and are similar to entities (and some relations) employed in Named Entity Recognition tasks, such as TREC and ACE.
- c. *Minor Syntactic Category Parsing*: e.g., locatives, purpose clauses, rationale clauses, temporal adjuncts.
- d. *Subphrasal Syntactic Cue Recognition*: e.g., genitives, partitives, bare plural/determiner distinctions, infinitivals, negatives.

The notion of a selection context pattern, as produced by a human annotator, is expressed as a BNF specification in Table 1.¹ This specification relies on word order to specify argument position, and is easily translated to a template with slots allocated for each argument. Within this grammar, a semantic role can be specified for each argument, but this information currently is not used in the automated processing.

English contains only about 8,000 verbs, of which we estimate that about 30% have only one basic pattern. The rest are evenly split between verbs having 2-3 patterns and verbs having more than 4 or

¹Round brackets indicate optional elements of the pattern, and curly brackets indicate syntactic constituents.

CPA-Pattern → Segment verb-lit Segment verb-lit Segment Segment verb-lit CPA-Pattern ';' Element	
Segment → Element Segment Segment '' Segment '' '(' Segment ')' Segment ' ' Segment	
Element → literal '[' Rstr ArgType ']' '[' Rstr literal ']' '[' Rstr ']' '[' NO Cue ']' '[' Cue ']'	
Rstr → POS Phrasal Rstr ' ' Rstr epsilon	
Cue → POS Phrasal AdvCue	
AdvCue → ADV '[' AdvType ']'	
AdvType → Manner Dir Location	
Phrasal → OBJ CLAUSE VP QUOTE	
POS → ADJ ADV DET POSDET COREF POSDET REFL-PRON NEG MASS PLURAL V INF PREP V-ING CARD QUANT CONJ	
ArgType → '[' SType ']' '[' SType '=' SubtypeSpec ']' ArgType ' ' ArgType ' ' '[' SType ArgIdx ']' '[' SType ArgIdx '=' SubtypeSpec ']'	
SType → AdvType TopType Entity Abstract PhysObj Institution Asset Location Human Animate Human Group Substance Unit of Measurement Quality Event State of Affairs Process	
SubtypeSpec → SubtypeSpec ' ' SubtypeSpec SubtypeSpec '&' SubtypeSpec Role Polarity LSet	
Role → Role Role ' ' Role Beneficiary Meronym Agent Payer	
Polarity → Negative Positive	
LSet → Worker Pilot Musician Competitor Hospital Injury Ailment Medicament Medical Procedure Hour-Measure Bargain Clothing BodyPart Text Sewage Part Computer Animal	
ArgIdx → <number>	verb-lit → <verb-word-form>
literal → word	word → <word>
CARD → <number>	NEG → not
POSDET → my your ...	INF → to
QUANT → CARD a lot longer more many ...	

Table 1: Pattern grammar

more patterns. About 20 light verbs have between 100 and 200 patterns each. This is less alarming than it sounds, because the majority of light verb patterns involve selection of just one specific nominal head, e.g., **take account**, **take plunge**, **take photograph**, with few if any alternations. The pattern sets for verbs of different frequency groups differ in terms of the number and type of features each pattern requires, the number of patterns in a set for a given verbs, the number of alternations for each pattern, and the type of selectional preferences affecting the verb's arguments.

Brandeis Shallow Ontology

The Brandeis Shallow Ontology (BSO) is a shallow hierarchy of types selected for their prevalence in manually identified selection context patterns. At the time of writing, there are just 65 types, in terms of which patterns for the first one hundred verbs have been analyzed. New types are added occasionally, but only when all possibilities of using existing types prove inadequate. Once the set of manually extracted patterns is sufficient, the type system will be re-populated and become pattern-driven.

The BSO type system allows multiple inheritance (e.g. Document \sqsubseteq PhysObj and Document \sqsubseteq Information). The types currently comprising the ontology are listed above. The BSO contains type assignments for 20,000 noun entries and 10,000 nominal collocation entries.

TOPTYPE	Location
Event	Dwelling
Action	Accommodation
<u>SpeechAct</u>	Energy
Activity	Abstract
Process	Attitude
State	Emotion
Entity	Responsibility
<u>PhysObj</u>	Privilege
Artifact	Rule
Machine	Information
Vehicle	Document
Hardware	Music
Document	Artwork
Music	Film
Artwork	Program
Film	Software
Program	Word
Software	Language
Medium	Concept
Garment	Property
Drug	<u>VisibleFeature</u>
Substance	Color
Vapor	Shape
Animate	<u>TimePeriod</u>
Bird	Holiday
Horse	<u>CourseOfStudy</u>
Person	Cost
Human Group	Asset
Plant	Route
<u>PlantPart</u>	
Body	
<u>BodyPart</u>	
Institution	
<u>HumanGroup</u>	

Corpus-driven Type System

The acquisition strategy for selectional preferences for predicates proceeds as follows:

- (5) a. Partition the corpus occurrences of a predicate according to the selection contexts pattern grammar, distinguished by the four levels of constraints mentioned in (4). These are uninterpreted patterns for the predicate.
- b. Within a given pattern, promote the statistically significant literal types from the corpus for each argument to the predicate. This induces an interpretation of the pattern, treating the promoted literal type as the specific binding of a shallow type from step (a) above.
- c. Within a given pattern, coerce all lexical heads in the same shallow type for an argument, into the promoted literal type, assigned in (b) above. This is a coercion of a lexical head to the interpretation of the promoted literal type induced from step (b) above.

In a sense, (5a) can be seen as a broad multi-level partitioning of the selectional behavior for a predicate according to a richer set of syntactic and semantic discriminants. Step (5b) can be seen as capturing the norms of usage in the corpus, while step (5c) is a way of modeling the exploitation of these norms in the language (through coercion, metonymy, and other generative operations). To illustrate the way in which CPA discriminates uninterpreted patterns from the corpus, we return to the verb *treat* as it is used in the BNC. Although there are three basic senses for this verb, the two major senses, as illustrated in (1) above, emerge as correlated with two distinct context patterns, using the discriminant constraints mentioned in (4) above. For the full specification for this verb, see www.cs.brandeis.edu/~arum/cpa/treat.html.

- (6) a. [[Person 1]] treat [[Person 2]]; NO [Adv[Manner]]
- b. [[Person 1]] treat [[Person 2]] [Adv[Manner]]

Given a distinct (contextual) basis on which to analyze the actual statistical distribution of the words in each argument position, we can promote statistically relevant and significant literal types for these positions. For example, for pattern (a) above, this induces **Doctor** as Person 1, and **Patient** as bound to Person 2. This produces the interpreted context pattern for this sense as shown below.

- (7) [[doctor]] treat [[patient]]

Promoted literal types are corpus-derived and predicate-dependent, and are syntactic heads of phrases that occur with the greatest frequency in argument positions for a given sense pattern; they are subsequently assumed to be subtypes of the particular shallow type in the pattern. Step (5c) above then

enables us to bind the other lexical heads in these positions as *coerced* forms of the promoted literal type. This can be seen below in the concordance sample, where *therapies* is interpreted as **Doctor**, and *people* and *girl* are interpreted as **Patient**.

- (8) a. returned with a doctor who **treated** the girl till an ambulance arrived.
- b. more than 90,000 people have been **treated** for cholera since the epidemic began
- c. nonsurgical therapies to **treat** the breast cancer, which may involve

Model Bias

The assumption within GL is that semantic types in the grammar map systematically to default syntactic templates (cf. Pustejovsky (1995)). These are termed *canonical syntactic forms (CSFs)*. For example, the CSF for the type **proposition** is a tensed S. There are, however, many possible realizations (such as infinitival S and NP) for this type due to the different possibilities available from generative devices in a grammar, such as coercion and co-composition. The resulting set of syntactic forms associated with a particular semantic type is called a *phrasal paradigm* for that type. The model bias provided by GL acts to guide the interpretation of purely statistically based measures (cf. Pustejovsky (2000)).

2.2 Automatic Recognition of Pattern Use

Essentially, this subtask is similar to the traditional supervised WSD problem. Its purpose is (1) to test the discriminatory power of CPA-derived feature-set, (2) to extend and refine the inventory of features captured by the CPA patterns, and (3) to allow for predicate-based argument groupings by classifying unseen instances. Extension and refinement of the inventory of features should involve feature induction, but at the moment this part has not been implemented. During the lexical discovery stage, lexical sets that fill some of the argument slots in the patterns are instantiated from the training examples. As more predicate-based lexical sets within shallow types are explored, the data will permit identification of the types of features that unite elements in lexical sets.

2.3 Automatic Pattern Acquisition

The algorithm for automatic pattern acquisition involves the following steps:

- (9) a. Collect all constituents in a particular argument position;
- b. Identify syntactic alternations;
- c. Perform clustering on all nouns that occur in a particular argument position of a given predicate;
- d. For each cluster, measure its relatedness to the known lexical sets, obtained previously during the lexical discovery stage and extended

through WSD of unseen instances. If none of the existing lexical sets pass the distance threshold, establish the cluster as a new lexical set, to be used in future pattern specification.

Step (9d) must include extensive filtering procedures to check for shared semantic features, looking for commonality between the members. That is, there must be some threshold overlap between subgroups of the candidate lexical set and the existing semantic classes. For instance, checking if, for a certain percentage of pairs in the candidate set, there already exists a set of which both elements are members.

3 Current Implementation

The CPA patterns are developed using the British National Corpus (BNC). The sorted instances are used as a training set for the supervised disambiguation. For the disambiguation task, each pattern is translated into into a set of preprocessing-specific features.

The BNC is preprocessed using the Robust Accurate Statistical Parsing system (RASP) and semantically tagged with BSO types. The RASP system (Briscoe and Carroll (2002)) tokenizes, POS-tags, and lemmatizes text, generating a forest of full parse trees for each sentence and associating a probability with each parse. For each parse, RASP produces a set of grammatical relations, specifying the relation type, the headword, and the dependent element. All our computations are performed over the single top-ranked tree for the sentences where a full parse was successfully obtained. Some of the grammatical relations identified by RASP are shown in (10).

- (10) **subjects:** ncsbj, clausal (csubj, xsubj)
objects: dobj, iobj, clausal complement
modifiers: adverbs, modifiers of event nominals

We use endocentric semantic typing, i.e., the headword of each constituent is used to establish its semantic type. The semantic tagging strategy is similar to the one described in Pustejovsky et al. (2002). Currently, a subset of 24 BSO types is used for semantic tagging.

A CPA pattern is translated into a feature set, which in the current implementation uses binary features. It is further complemented with other discriminant context features which, rather than distinguishing a particular pattern, are merely *likely* to occur with a given subset of patterns; that is, the features that only partially determine or co-determine a sense. In the future, these should be learned from the training set through feature induction from the training sample, but at the moment, they are added manually. The resulting feature matrix for each pattern contains features such as those in (11) below. Each pattern is translated into a template of 15-25 features.

- (11) **Selected context features:**
- a. **obj_institution:** object belongs to the BSO type 'Institution'
 - b. **subj_human_group:** subject belongs to the BSO type 'HumanGroup'
 - c. **mod_adv_ly:** target verb has an adverbial modifier, with a -ly adverb
 - d. **clausal_like:** target verb has a clausal argument introduced by 'like'
 - e. **iobj_with:** target verb has an indirect object by 'with'
 - f. **obj_PRP:** direct object is a personal pronoun
 - g. **stem_VVG:** the target verb stem is an -ing form

Each feature may be realized by a number of RASP relations. For instance, a feature dealing with objects would take into account RASP relations 'dobj', 'obj2', and 'ncsubj' (for passives). The features such as (11a)-(11e) are typically taken directly from the pattern specification, while features such as in (11f) and (11g) would typically be added as co-determining the pattern.

4 Results and Discussion

The experimental trials performed to date are too preliminary to validate the methodology outlined above in general terms for the WSD task. Our results are encouraging however, and comparable to the best performing systems reported from Senseval 2. For our experiments, we implemented two machine learning algorithms, instance-based k-Nearest Neighbor, and a decision tree algorithm (a version of ID3). For these experiments, kNN was run with the full training set. Table 2 shows the results on a subset of verbs that have been processed, also listing the number of patterns in the pattern set for each of the verbs.²

verb	number of patterns	training set	accuracy	
			ID3	kNN
edit	2	100	87%	86%
treat	4	200	45%	52%
submit	4	100	59%	64%

Table 2: Accuracy of pattern identification

Further experimentation is obviously needed to adequately gauge the effectiveness of the selection context approach for WSD and other NLP tasks. It is already clear, however, that the traditional sense enumeration approach, where senses are associated with individual lexical items, must give way to a model where senses are assigned to the contexts within which words appear. Furthermore, because the variability of the stereotypical syntagmatic patterns that are associated with words appears to be relatively small, such information can be encoded as

²Test set size for each lemma is 100 instances, selected out of several randomly chosen segments of BNC, non-overlapping with the training set

lexically-indexed contexts. A comprehensive dictionary of such contexts could prove to be a powerful tool for a variety of NLP tasks.

References

- S. Abney and M. Light. 1999. Hiding a semantic hierarchy in a markov model.
- E. Agirre and D. Martinez. 2001. Learning class-to-class selectional preferences. In Walter Daelemans and Rémi Zajac, editors, *Proceedings of CoNLL-2001*, pages 15–22. Toulouse, France.
- E. Agirre, D. Martinez, and L. Marquez. 2002. Syntactic features for high precision word sense disambiguation. *COLING 2002*.
- S. Atkins, C. Fillmore, and C. Johnson. 2003a. Lexicographic relevance: Selecting information from corpus evidence. *International Journal of Lexicography*, 16(3):251–280, September.
- S. Atkins, M. Rundell, and H. Sato. 2003b. The contribution of Framenet to practical lexicography. *International Journal of Lexicography*, 16(3):333–357, September.
- C. Baker, C. Fillmore, and B. Cronin. 2003. The structure of the Framenet database. *International Journal of Lexicography*, 16(3):281–296, September.
- T. Briscoe and J. Carroll. 1997. Automatic extraction of subcategorization from corpora. *Proceedings of the 5th ANLP Conference, Washington DC*, pages 356–363.
- T. Briscoe and J. Carroll. 2002. Robust accurate statistical annotation of general text. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002), Las Palmas, Canary Islands, May 2002*, pages 1499–1504.
- J. Carroll and D. McCarthy. 2000. Word sense disambiguation using automatically acquired verbal preferences.
- M. Ciaramita and M. Johnson. 2000. Explaining away ambiguity: Learning verb selectional preference with Bayesian networks. *COLING 2000*.
- S. Clark and D. Weir. 2001. Class-based probability estimation using a semantic hierarchy. *Proceedings of the 2nd Conference of the North American Chapter of the ACL. Pittsburgh, PA*.
- C. Fillmore, C. Johnson, and M. Petruck. 2003a. Background to Framenet. *International Journal of Lexicography*, 16(3):235–250, September.
- C. Fillmore, M. Petruck, J. Ruppenhofer, and A. Wright. 2003b. Framenet in action: The case of attaching. *International Journal of Lexicography*, 16(3):297–332, September.
- C. Fillmore. 1975. *Santa Cruz Lectures on Deixis*. Indiana University Linguistics Club. Bloomington, IN.
- M. A. K. Halliday. 1966. Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins, editors, *In Memory of J. R. Firth*. Longman.
- P. Hanks. 2004. The syntagmatics of metaphor (forthcoming). *International Journal of Lexicography*, 17(3), September. forthcoming.
- A. Korhonen and J. Preiss. 2003. Improving subcategorization acquisition using word sense disambiguation. *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics. Sapporo, Japan*.
- A. Korhonen. 2002. Subcategorization Acquisition. *PhD thesis published as Technical Report UCAM-CL-TR-530. Computer Laboratory, University of Cambridge*.
- M. Light and W. Greiff. 2002. Statistical models for the induction and use of selectional preferences. *Cognitive Science, Volume 26(3)*, pp. 269–281.
- D. McCarthy. 1997. Word sense disambiguation for acquisition of selectional preferences. In Piek Vossen, Geert Adriaens, Nicoletta Calzolari, Antonio Sanfilippo, and Yorick Wilks, editors, *Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 52–60. Association for Computational Linguistics, New Brunswick, New Jersey.
- M. Palmer, H. T. Dang, and C. Fellbaum. 2004. Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering. Preprint*.
- J. Preiss and D. Yarowsky, editors. 2001. *Proceedings of the Second Int. Workshop on Evaluating WSD Systems (Senseval 2)*. ACL2002/EACL2001.
- J. Pustejovsky and P. Hanks. 2001. Very Large Lexical Databases: A tutorial. *ACL Workshop, Toulouse, France*.
- J. Pustejovsky, A. Rumshisky, and J. Castano. 2002. Rerendering Semantic Ontologies: Automatic Extensions to UMLS through Corpus Analytics. In *LREC 2002 Workshop on Ontologies and Lexical Knowledge Bases. Las Palmas, Canary Islands, Spain*.
- J. Pustejovsky. 1995. *Generative Lexicon*. Cambridge (Mass.): MIT Press.
- J. Pustejovsky. 2000. Lexical shadowing and argument closure. In Y. Ravin and C. Leacock, editors, *Lexical Semantics*. Oxford University Press.
- P. Resnik. 1996. Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, 61:127–159.
- M. Rooth, S. Riezler, D. Prescher, G. Carroll, and F. Beil. 1999. Inducing a semantically annotated lexicon via EM-based clustering. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL'99), Maryland*.
- J. Sinclair, P. Hanks, and et al. 1987. *The Collins Cobuild English Language Dictionary*. HarperCollins, 4th (2003) edition. Published as Collins Cobuild Advanced Learner's English Dictionary.
- J. M. Sinclair. 1991. *Corpus, Concordance, Collocation*. Oxford University Press.
- J. Stetina, S. Kurohashi, and M. Nagao. 1998. General word sense disambiguation method based on A full sentential context. In Sanda Harabagiu, editor, *Use of WordNet in Natural Language Processing Systems: Proceedings of the Conference*, pages 1–8. Association for Computational Linguistics, Somerset, New Jersey.
- M. Stevenson and Y. Wilks. 2001. The interaction of knowledge sources in word sense disambiguation. *Computational Linguistics*, 27(3), September.
- K. Yamashita, K. Yoshida, and Y. Itoh. 2003. Word sense disambiguation using pairwise alignment. *ACL2003*.
- D. Yarowsky. 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. *Proc. COLING92, Nantes, France*.
- D. Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Meeting of the Association for Computational Linguistics*, pages 189–196.