

[Sabine Schulte im Walde: "The induction of verb frames and verb classes from corpora." To appear as chapter 61 in: Anke Lüdeling and Merja Kytö (eds): *Corpus Linguistics. An International Handbook*. Mouton de Gruyter, Berlin.]

5 61 The induction of verb frames and verb classes from corpora

10 Creating lexical information resources manually is an expensive effort: It takes a long time to define detailed lexical knowledge, then the information needs to be updated regularly because of neologisms, sublanguages and language change, and the lexicon will rarely if ever be complete. For these reasons and also given the increasing availability of computing power and corpus resources, one line of research at the interface of corpus and computational linguistics aims at an automatic acquisition of lexical information, utilising existing corpora and applying computational algorithms. The retrieved lexical information is stored in machine-readable lexicons, and can be updated dynamically and quickly. Also, the resulting lexical resources can be integrated into computational tasks and applications in Natural Language Processing (NLP), such as parsing, machine translation, question answering, and many more.

20 Within the area of automatic lexical acquisition, the induction of lexical verb information has been a major focus, because verbs play a central role for the structure and the meaning of sentences and discourse. The levels of information that are relevant for a verb lexicon concern all lexical aspects of verbs, ranging from phonological and morphological to syntactic, semantic, and pragmatic criteria. This article introduces work that focuses on the acquisition of lexical verb properties at the syntax-semantics interface, and addresses the automatic induction of verb frames from corpora (section 61.1), and the acquisition of verb classes (section 61.2). As it is true for automatic lexical acquisition in general, the combination of corpus data and computational algorithms is not always straightforward, and we find a variety of solutions: corpus data can be used on various levels of annotation, i.e., as raw text, with part-of-speech tags, as parsed text with structural information, etc. (cf. articles 24-34 on preprocessing corpora); algorithms might fit better or worse to a certain acquisition task, depending on the mathematical properties of the algorithm in relation to the linguistic task (cf. article 38); the acquisition results are often difficult to compare because they rely on different theoretical assumptions and produce different types of output, and there is not always an evaluation method available. In the course of this article, each section provides an overview of existing approaches to the induction of verb frames and verb classes, and describes their assumptions, procedures and evaluations.

35 61.1 Induction of verb frames from corpora

The potential of verbs to choose their complements¹ is referred to as 'verb subcategorisation' or 'verb valency', and a combination of functional complements that are evoked by a verb is often called a 'verb subcategorisation frame', or simply a 'verb frame'. For example, the verb 'bake' can subcategorise for a direct object (in addition to the obligatory subject), as in (i).

(i) Elsa bakes a chocolate cake.

Alternatively, 'bake' might subcategorise for a direct object plus an indirect object, or a temporal prepositional phrase, as illustrated by (ii) and (iii), but cannot be used with e.g. a *that*-clause as in (iv). With a different verb such as 'say' the frame in (iv) would have been acceptable: 'Elsa says that she likes cakes'.

(ii) Elsa bakes Tim a chocolate cake.

(iii) The chocolate cake baked for 1 hour.

(iv) *Elsa bakes that she likes cakes.

Verb frames might distinguish between obligatory and optional verb complements; however, this distinction is not always clear-cut, cf. the prepositional phrase (PP) in example (iii): is the PP obligatory or optional? (See Meyers/Macleod/Grishman (1994) as one definition of the criteria for distinguishing between obligatory and optional verb complements.)

Typically, verb frames are illustrated as a set of the complements they include, such as $\{Subj, Obj-dir, Obj-indir\}$, or $\{Subj, PP-for\}$. Depending on the framework, the details of the frame description vary. For example, languages where subjects are obligatory need not explicitly include subjects into the verb frame; languages with case marking tend to refer to the case of noun phrases instead of using direct vs. indirect

¹ In this article, the term 'complement' is used to subsume the terms 'argument' and 'adjunct'.

objects; some approaches might distinguish PP arguments and adjuncts, others not; PPs can be referred to by a very general label (such as `PP` only), by semantic category labels (such as `PP-tmp`, `PP-loc`), or by the specific preposition (such as `PP-for`, `PP-at`); etc.

60 Subcategorisation/valency is not restricted to the syntactic options of verb complements, but also refers to the semantic and the pragmatic level, cf. Helbig (1992) and Fischer (1999), among others. Example (v) presents a clause where `bake` subcategorises for a direct object as in (i), but appears strange, because a stone is typically not baked. The degree of acceptability with respect to the semantic realisation of a complement varies, so a verb is said to define `selectional preferences` for its complements.

65 (v) ?Elsa bakes a stone.

Selectional preferences do not only refer to the syntactic function of a complement within a verb frame, but take into account the `semantic role` of the respective complement, cf. the thematic proto-roles in Dowty (1991), and the argument structure in Grimshaw (1992). For example, the direct object in the causative transitive clause `Elsa melts the chocolate` and the subject in the inchoative intransitive variant `The chocolate melted` is the nominal phrase `the chocolate`, and this NP represents the patient role of the verb `bake` in both variants. As the example illustrates, selectional preferences are required by the semantic roles of complements, which are in turn determined by the verb and the syntactic function of the complement within a certain verb frame. The phenomenon that verb frames and semantic roles can be used in alternative constructions is referred to as `diathesis alternation`, cf. Levin (1993) for a prominent collection of English verb alternations.

The induction of verb frames from corpora was one of the first issues when empirical lexical acquisition from corpora started out. The reason for this specific interest is that subcategorisation frames of verbs provide useful information for the structural analysis of sentences, which is necessary in e.g. parsing, cf. chapter 29. For example, Briscoe/Carroll (1993) found that half of parse failures on unseen data are caused by inaccurate subcategorisation information. Later, the syntactic frame information in automatic lexical acquisition was gradually expanded to include semantic information referring to selectional preferences or semantic roles, and also definitions of diathesis alternations. In what follows in this section is organised accordingly: section 61.1.1 describes approaches to acquiring syntactic verb frame types, and section 61.1.2 introduces extensions to the syntactic frame definitions.

61.1.1 Approaches to inducing subcategorisation frames

90 Approaches to automatically acquiring lexical verb information on subcategorisation frames can be defined and distinguished with respect to several dimensions, which we specify here as (a) to (e):

(a) *Corpus selection and preparation*: Which corpus is selected as the data resource, and what kind of annotation is provided?

95

(b) *Frame types*: How many and which types of verb frames are distinguished?

(c) *Acquisition method*: Which computational methods are used, in order to induce the subcategorisation frames?

100

(d) *Filtering*: Are the subcategorisation frames as obtained under (c) filtered for noise, and what kind of method is used for filtering?

(e) *Evaluation*: How is the resulting frame information evaluated?

105

In what follows, this section elaborates on the above criteria and exemplifies them by approaches to subcategorisation acquisition. The examples refer to representative (but not exhaustive) work for English; additional approaches for languages other than English follow.

110

Corpus Selection:

The first step (a) in frame acquisition is to select a corpus (cf. article 22 for an overview of corpora). As is true for empirical acquisition in general, researchers try to use as much data as are available (with respect to the language they are concerned with) and which can be processed with applicable computing resources. For

115 example, in the early stages of subcategorisation acquisition, Brent (1993) used 2.6 million words of the
 Wall Street Journal (WSJ), Ushioda et al. (1993) used 600.000 words of the same corpus, Manning (1993)
 used 4 million words of the New York Times newswire, and Briscoe/Carroll (1993) used 1.2 million words
 from the Susanne corpus, the Corpus of Spoken English (SEC), and the Lancaster-Oslo/Bergen Corpus
 (LOB). Comparing the early work with more recent approaches illustrates the increasing amount of available
 120 data and the decreasing restrictions on data processing. For example, Carroll/Rooth (1998) used the whole
 British National Corpus (BNC) with 117 million words.

In addition to the quantitative influence of the corpus size, the acquisition result is determined by the
 qualitative properties of the corpus, i.e., the genre of the corpus, the speech type (written vs. spoken corpora),
 the corpus age, etc. As a prominent example highlighting the influence of corpus choice on acquisition
 results, Roland/Jurafsky (2002) compared the frequencies of verb subcategorisation frames as obtained from
 125 five different corpora; two corpora were derived from psychological experiments in which participants were
 asked to produce single isolated sentences; two corpora were written texts, extracted from the Brown corpus
 and the Wall Street Journal corpus (Marcus/Marcinkiewicz/Santorini 1993), and one corpus contained
 telephone conversations. Roland/Jurafsky reported differences between the frame types and the frame type
 frequencies, and that the two major sources of the acquisition differences were (a) the discourse type, and (b)
 130 the semantic choices, i.e., the word senses represented in the corpora.

Corpus Annotation:

The methods for frame induction differ in the level of annotation they presuppose. Some approaches use raw
 corpus data, others either preprocess the corpus (i.e., lemmatiser/tagger/parser are applied to annotate the raw
 135 data), or use existing annotations provided by the corpus, such as the WSJ which is manually annotated on
 several levels (i.e., lemmas, part-of-speech (POS) tags, parse trees). For example, Brent (1993) performed
 frame acquisition from raw corpus data; Ushioda et al. (1993) used corpus data annotated with part-of-
 speech tags, and Manning (1993), Briscoe/Carroll (1997) as well as most subsequent work assumed partially
 or fully parsed corpus data. Recent work such as Kinyon/Prolo (2002) and O'Donovan et al. (2005) relied on
 140 the annotation provided by a treebank (here: the Penn Treebank). Approaches which work on unannotated
 data are a priori more restricted in the linguistic details of their frame variants than approaches which take
 deeper morpho-syntactic information into account.

Frame Types:

Existing approaches differ strongly with respect to the desired number and types of subcategorisation frames
 they induce: Are all available subcategorisation frames relevant, or are the frame types restricted to a subset?
 How fine-grained are the frame types, e.g. do they distinguish between different kinds of clauses or
 prepositional phrases? And do the approaches address the distinction between arguments and adjuncts, or
 generalise over the functions? For example, Brent's approach (1993) detected six frame types which only
 150 addressed direct objects and subcategorised clauses and infinitives. Ushioda et al. (1993) used a larger
 variety of complement types, but also restricted the experiments to six frame types, not distinguishing
 arguments and adjuncts. Manning (1993) defined 19 frame types with limited information on prepositions,
 still not distinguishing between arguments and adjuncts. Briscoe/Carroll (1997) acquired lexical information
 on 163 frame types, including the distinction between arguments and adjuncts, and a fine-grained reference
 155 to prepositional phrase types. Carroll/Rooth (1998) allowed all combinations of verb-adjacent constituents as
 frame types, as based on their context-free grammar parses.

Approaches which start out with no or few restrictions on frame types are more challenging (because they do
 not rely on existing frame information) but also more flexible than those which induce the syntax-semantics
 structure from a treebank. However, both classes of approaches are important; the more flexible approaches
 160 enable the induction of non-pre-existing, domain-independent subcategorisation lexicons and allow for
 unforeseen categories, and the treebank-based approaches are strong for theory-related, domain-specific
 subcategorisation knowledge.

As the overview of selected approaches illustrates, the degree of details within the frame types and – related
 to this – the number of types depend on the underlying corpus information and also on individual decisions
 165 of which complements to include in the frames. It is important to note that there is no optimum with respect
 to size and details of the frame types. The 'optimal' subcategorisation lexicon depends on the NLP
 task/application which uses such a lexicon. For example, the argument/adjunct distinction is more relevant in
 machine translation (where it is important to relate the constituent functions between two languages in some
 detail) than in question answering (where the question/answer functions are rather generalised to enhance the
 170 query results).

Acquisition Method:

Another criterion in frame induction refers to the method for the lexical acquisition. Here we need to distinguish two steps, which might be interrelated: (i) the identification of the verbs in the corpus, and (ii) the identification and quantification of the frame types. The identification of the verbs is more or less difficult with respect to the level of corpus annotation that is accessed: raw corpus data provides less cues than part-of-speech tagged or even parsed corpus data. Also, languages with richer morphology (such as German) facilitate the detection of verbs, as compared to languages like English with poor morphology. Below, a series of methods are introduced, whose chronological order illustrates an increase in both the amount of corpus annotation and also the complexity of the acquisition approach.

For example, Brent (1993) identified English verbs in a raw corpus as all lexical items that appeared both with and without the suffix ‘-ing’. The result was filtered by heuristics on the lexical context, e.g. potential verbs which directly follow a determiner were not considered. The verb complements were identified by a finite-state grammar which defined linear patterns, such as ‘to V’ referring to a subcategorised infinitival clause. For its simplicity, Brent’s approach is surprisingly successful, but cannot be extended to sufficiently cover additional frame types, as no reliable cues exist for many frames.

Ushioda et al. (1993) used part-of-speech tags to identify verbs; to identify frame types, they defined a finite-state grammar for chunking, and regular expressions for linear chunk patterns. As Brent’s approach above, the procedure is sufficient for a small number of frame types, but difficult to extend.

Manning (1993) used a finite-state parser to parse only clauses in the corpus with auxiliaries; relying on the restricted sentence structure, he identified the constituents following the verb as the verb complements. His approach is more reliable for a larger set of frame types, but restricts itself to a certain surface pattern, i.e., clauses with auxiliaries.

Later approaches made use of more complex corpus annotation: Briscoe/Carroll (1997) used the ranked output analyses of a probabilistic parser trained on a treebank, and extracted the verb and the subcategorised complements (plus their lexical heads) from the parses. Carroll/Rooth (1998) used a head-lexicalised probabilistic context-free grammar (HL-PCFG), trained the grammar with an unsupervised algorithm, and induced lexicalised subcategorisation information from the trained grammar model. The approaches by Briscoe and Carroll as well as Carroll and Rooth allowed all patterns which occurred according to their grammar, and derived the frame types according to these patterns.

Kinyon/Prolo (2002) defined a mapping from the Penn Treebank annotation to obtain verbs and their subcategorisation frame types. O’Donovan et al. (2005) derived their verb and frame information after they had performed an automatic annotation of the Penn Treebank with LFG (Lexical Functional Grammar) structures. The latter two approaches therefore relied on the definitions in a treebank to induce the verb-frame lexicon.

Filtering:

Once the verb frame information is acquired, most approaches perform an additional step: they filter the empirical outcome. Brent (1993) suggested a hypothesis test to determine the reliable association between a verb and a frame type, referring to a binomial test. His filter was adapted in subsequent work, e.g., by Manning (1993) and Briscoe/Carroll (1997). However, Korhonen/Gorrell/McCarthy (2000) showed that a much simpler filter, which defines a cut-off on the relative frequency of a verb-frame pair, is sufficient and performs better than the hypothesis test. Following a different intuition, Korhonen (2002) – whose work built on Briscoe/Carroll (1997) – suggested a filter based on verb semantics: she demonstrated the usefulness of semantic verb classes (cf. section 61.2) by smoothing the empirical subcategorisation frames with back-off estimates on the verbs’ semantic classes by Levin (1993), and subsequently applied a simple threshold to the estimates. The range of approaches to filtering illustrates that, on the one hand filtering is seen as an important step after the acquisition procedure, but on the other hand the type and the complexity of the filtering methods differ strongly.

Evaluation:

There are multiple possibilities for evaluating the empirical frame information. Existing approaches performed either (a) manual judgement, or (b) an evaluation against frame types listed in existing manually built lexicons, or (c) an evaluation by integrating the frame information into an NLP task or application. For example, Brent (1993) evaluated his English subcategorisation frames by hand judgement, reporting an f-score of 73.85%. Wauschkuhn (1999) did the same for German on a choice of seven verbs. He reported an f-score of 61.86%. Manning (1993) and Carroll/Rooth (1998) evaluated their frames against the ‘Oxford Advanced Learner’s Dictionary’ (Hornby 1985) and reported an f-score of 58.20% and 76.95%, respectively. Briscoe/Carroll (1997) used the Alvey NL Tools Dictionary (Boguraev et al. 1987) and the COMLEX

230 Syntax Dictionary (Grishman/Macleod/Meyers 1994) and achieved 46.09% f-score. O'Donovan et al. (2005) also used COMLEX, achieving 27.3%/64.3% with/without PP specifications. Schulte im Walde (2002a, 2002b) used the German dictionary 'Duden – Das Stilwörterbuch' (Bibliographisches Institut & F. A. Brockhaus AG), reporting an f-score of 57.24%/62.30% with/without PP specifications.

235 Even though the f-scores above suggest qualitative differences between the various approaches, it is important to note that a comparison between the results is difficult. Generally speaking, a manual inspection of the frame results has the advantage of being more flexible than comparing the frame information against pre-defined dictionary entries. However, in case the manual judgement is performed by only one person, it runs into the danger of being subjective, so it is necessary to rely on multiple annotators for evaluation. On
 240 the other hand, an evaluation against dictionaries is bound to the assumptions and definitions in a dictionary, which might differ from those in the lexical acquisition approach: the dictionary provides more details in some cases but less details in others, with respect to the kinds of complements included in the frame types (such as the subcategorisation of clauses, or arguments vs. adjuncts), and the granularity of the complement properties (such as the kinds of prepositional phrases, or case information in morphologically rich
 245 languages). Last but not least, the automatic induction of verb frames is more or less difficult with regard to how many different frame types the approaches call for, and how fine-grained the frame type information is. In conclusion, one should only compare the outcome of approaches whose target frame types are sufficiently similar, and whose evaluation methods are comparable to a large extent. The f-score results reported above should therefore be interpreted with caution.

250 Integrating a subcategorisation lexicon within an application is one way to compare the outcomes of various approaches, because the improved performance of the application when using subcategorisation information can be measured. This method has rarely been employed for comparison reasons, because the induced frame lexicons usually differ with respect to their original purpose, and also a comparison across languages is
 255 difficult. However, a few research groups did apply their result to NLP tasks or applications. Work based on Briscoe/Carroll (1997,2002) applied subcategorisation information to improve the coverage and accuracy of parsing systems: Carroll/Minnen/Briscoe (1998) used the 1997-system and showed that subcategorisation information significantly improved the accuracy of their wide-coverage parser in inducing grammatical relations. Carroll/Fang (2004) extended the 2002-system with a subcategorisation lexicon and showed that
 260 the extension helped a deep HPSG parser to improve its coverage and the parsing success rate.

Languages:

The above criteria and approaches illustrate the variety of frame acquisition ideas, and also the development over time, utilising an increasing amount of data and defining more complex algorithms and filters. So far,
 265 we have mainly referred to approaches for English. But we also find approaches to the automatic induction of syntactic frames in other languages than English: For German, Eckle (1999) performed a semi-automatic acquisition of subcategorisation information for 6,305 verbs. She worked on POS-tagged corpus data and defined linguistic heuristics by regular expression queries over the usage of 244 frame types including PP definitions. Wauschkuhn (1999) constructed a valency dictionary for 1,044 German verbs. He extracted a
 270 maximum of 2,000 example sentences for each verb from annotated corpus data, and constructed a context-free grammar for partial parsing. The syntactic analyses provided valency patterns, which were grouped in order to extract the most frequent pattern combinations, resulting in a verb-frame lexicon with 42 frame types. Schulte im Walde (2002a) developed a German context-free grammar containing frame-predicting grammar rules, and used the unsupervised training environment of HL-PCFGs (Carroll/Rooth 1998) to train
 275 the grammar on 18.7 million words of German newspaper corpora. She induced subcategorisation frames for more than 14,000 German verbs, for 38 purely syntactic frame types and a refinement of 178 frame types including prepositional phrase distinctions. For Portuguese and Greek, de Lima (2002) and Georgala (2003), respectively, also utilised the same HL-PCFG framework to learn verb-frame combinations for the respective languages. For Czech, Sarkar/Zeman (2000) used the syntactic dependency definitions in the Prague
 280 Dependency Treebank (PDT) to induce subcategorisation frames. A frame was defined as a subset of the annotated dependents of a verb in the treebank. As a major task, they learned the argument-adjunct distinction in the frame types. For Dutch, Spranger/Heid (2003) developed a chunker to extract verb subcategorisation. For French, Chesley/Salmon-Alt (2006) created a multi-genre corpus with random occurrences of 104 frequent verbs from the Frantext online literary database. They applied a dependency-
 285 parser to the corpus and obtained 27 different subcategorisation frames as any combination of a restricted set of constituents (direct objects, pre-specified PPs, clauses, adjectival phrases, and reflexive clitic NPs).

In addition to the `usual` differences between the approaches to subcategorisation acquisition (cf. the discussions in (a) to (e) above), the approaches across languages naturally differ as a consequence of the properties of the respective languages, such as morphological marking, word order, etc. For example, Carroll/Root (1998) defined a HL-PCFG with flat grammar rules for English, whereas Schulte im Walde (2002a) – who used the same framework for German – defined mostly binary context-free rules because the freer word order in German would have required an enormous amount of flat rules (covering all possible constellations of complement orderings, combined with adverbial modifiers, etc.), creating training problems for a lexicalised grammar, and possibly causing a sparse-data problem. Another example of language differences concerns the relevance of complement types within the verb frames. For example, in some languages subjects are obligatory (e.g., English, German), whereas in others (e.g., Italian, Spanish) they are not, so the relevance for including subject information into the frames differs; also, certain complements such as adjectival phrases are of minor (e.g., in German) vs. major (e.g., in French) importance with respect to their productivity.

61.1.2 Approaches to empirical extensions of verb frames

So far, we have discussed the induction of purely syntactic verb frames, plus refinement by prepositional phrase types in some approaches. But syntactic frames are only one part of verb subcategorisation, as mentioned above. In this section, we address the acquisition of verb frames with additional semantic subcategorisation, i.e., we introduce approaches which empirically define selectional preferences or semantic roles for verb frames. In addition, we refer to approaches which build on the induction of syntactic and semantic subcategorisation, and address the diathesis alternation of verbs, namely the alternative usage of frames and roles.

Selectional Preferences:

As demonstrated in section 61.1 of this article, the degree of acceptability with respect to the semantic realisation of a complement varies, so a verb is said to define `selectional preferences` for its complements. From a practical point of view, selectional preferences for complements are useful because they refer to a generalisation of specific complement heads and therefore improve a sparse-data situation. For example, in lexicalised parsing, the lexical heads that are incorporated into the parser cause a sparse-data problem. Referring to selectional preferences instead of specific lexical heads might help the parser because it is confronted with e.g. `drink a beverage` and can abstract from seen instances such as `drink tea`, `drink coffee`, etc. to unseen instances such as `drink cacao`.

In order to define selectional preferences for frame complements, it is necessary to refer to an inventory of semantic categories, such as `animate` vs. `inanimate`, or `banana` vs. `teacher`, etc. The choice of semantic categories and the level of granularity depend on the theoretical assumptions of the researchers, and in practice the categories are often restricted to the definitions in an existing resource. The reason is that, on the one hand we demand a generalisation over nominal complements in order to talk about abstract preferences, but on the other hand we do not a priori find generalisations in corpus data. So it is helpful to refer to an external categorisation. The example approaches to follow utilise `WordNet` (Fellbaum 1998), a lexical semantic taxonomy originally developed for English at the University of Princeton, and since then transferred to additional languages, cf. the Global WordNet Association (www.globalwordnet.org) for more details. The lexical database was inspired by psycholinguistic research on human lexical memory. It organises nouns, verbs, adjectives and adverbs into classes of synonyms (`synsets`). Words with several senses are assigned to multiple classes. The synsets are connected by lexical and conceptual relations such as hypernymy, hyponymy, meronymy, etc. The hypernym-hyponym relation imposes a multi-level hierarchical structure on the taxonomy. The noun synsets in WordNet provide a choice of semantic categories on different levels of generalisation, which can be used to define selectional preferences for verbs. For example, the verb `drink` would specify a strong selectional preference for the WordNet synset `beverage` with respect to its direct object, intuitively because on the one hand the synset generalises over its hyponyms such as `coffee`, `tea`, `milk`, etc., and on the other hand it is more specific to the verb's complements than its hypernyms such as `food`, or even `substance`.

In the following, a choice of approaches which utilise WordNet is presented. As described above, WordNet provides a framework that is suitable for defining selectional preferences, and has therefore be used

extensively for this task. The selection of approaches is by far not exhaustive, but provides an overview and pointers to more information on selectional preference acquisition, with and without WordNet.

350 Resnik (1997) defined selectional preference as the association strength between a predicate and the semantic categories of its complements. The starting point of his approach was co-occurrence counts of predicates and complements within a specific syntactic relationship (such as `direct object`), as based on a corpus. The co-occurrence counts were assigned to those WordNet synsets which contain the respective heads of the complements, and propagated upwards in the WordNet hierarchy. For ambiguous complements, the count was split over all WordNet synsets containing that complement. This procedure was repeated for all complements, and the counts were accumulated for each synset. Furthermore, the procedure was applied
355 twice: (a) for each specific predicate of interest, e.g., for specific verbs, and (b) without relation to a specific predicate, i.e., accumulating over a class of predicates such as all verbs in the corpus. The association strength was then calculated by applying the information-theoretic measure of relative entropy to the two probability distributions based on the complement counts over WordNet synsets: The *prior* probability of a complement class (i.e., a WordNet synset such as `beverage`) *regardless* of the identity of the predicate is compared with the *posterior* probability of a complement class *with regard* to a specific predicate. Relative entropy calculates the distance between the respective probability distributions; the more similar the two probability distributions are, the weaker the association between predicate and complement class, and therefore the weaker the selectional preference of the predicate for that class.

360 Li/Abe (1998) also based their approach on co-occurrence counts of predicates and complements within a specific syntactic relationship. The selectional preferences for a predicate-complement structure were described by a cut in the WordNet hierarchy, i.e., a set of WordNet nodes. The cut was determined by the Minimum Description Length (MDL), a principle from information theory for data compression and statistical estimation. A selectional preference model where the chosen set of WordNet nodes is nearer the WordNet root is simpler to describe (by means of the number of bits for encoding the model) but with poorer fit to the data, i.e., the specific WordNet leaves; a model nearer the WordNet leaves is more complex but with better fit to the data. The MDL principle finds the cut in the hierarchy which minimises the sum of encoding both the model and the data.
365

370 Abney/Light (1999) provided a stochastic generation model to determine the selectional preferences of a predicate-complement relationship. The co-occurrence probabilities were estimated by a Hidden Markov Model (HMM) for each predicate structure. The HMM was defined and trained on the WordNet hierarchy, with the initial state being the (artificial) root node of WordNet. Each HMM run was a path through the hierarchy from the root to a word sense, plus the word generated from the word sense. The most likely path indicated the verbs' selectional preferences.
375

380 Clark/Weir (2002) estimated the joint frequencies for a predicate-complement relationship and a specific WordNet class in the same way as Resnik (1997). Their generalisation procedure then used the statistical chi square test to find the most suitable class: A bottom-up check of each node in the WordNet hierarchy determined whether the probability of the parent class was significantly different from that of the children classes. In that case, the search was stopped at the respective child node as the most suitable selectional preference representation.
385

Even though the above approaches used different algorithms to calculate selectional preferences, they all rely on similar data (verb-complement co-occurrence counts from a chunker or a parser) and attempt to characterise the selectional preferences of a verb by WordNet noun synsets. A priori, it is difficult to tell whether any of the approaches is optimal. Brockmann/Lapata (2003) therefore compared the approaches by
390 Resnik, Li/Abe, and Clark/Weir with respect to German verbs and their NP and PP complements, using a common corpus. The models, as well as a combination of the models, were evaluated against human ratings, demonstrating that there was no method which performs best overall. They added a model combination, using multiple linear regression, and the combined method actually obtained a better fit with the experimental data than the single methods. The comparison demonstrates that it is not necessarily the case
395 that one approach outperforms all other approaches. Rather, it is important to compare the variety of approaches with respect to a certain task, or even try to find combinations that complement each other.

Semantic Roles:

400 A second strand of adding semantic information to subcategorisation frames is concerned with the definition of semantic roles for complements. Differently to selectional preferences, semantic roles are not generalisations of lexical heads, but represent the semantic relationship between a predicate and a complement within a certain frame type. To refer back to our example in section 61.1, the NP `the chocolate`

405 represents the patient role of the direct object in the transitive clause `Elsa melts the chocolate` and also of the subject in the inchoative intransitive variant `The chocolate melted`. In practical terms, semantic roles are useful in applications such as question answering, where e.g. a question word such as `who` in `who killed ...` needs to be matched to an agent role for the verb `kill`, abstracting over syntactic functions and lexical heads.

410 As for selectional preference acquisition, we do not a priori find semantic roles in corpus data; thus, the approaches to semantic role labeling attempt to induce regularities from unlabeled data, or rely on manually annotated data. In the following, two prominent projects concerned with semantic role labeling are introduced, `FrameNet` and `PropBank`. Within the projects, corpora are annotated with semantic information; the annotation is partly manual and partly semi-automatic; the semi-automatic labeling explores
415 unsupervised methods for role labeling. The annotated data can be used for supervised approaches to learning semantic subcategorisation.

FrameNet (Baker/Fillmore/Lowe 1998) is based on Fillmore's frame semantics (Fillmore 1982) and thus describes `frames`, i.e., the background and situational knowledge needed for understanding a word or expression. Each FrameNet frame provides its set of semantic roles, the participants and properties of the
420 prototypical situation. The Berkeley FrameNet project is building a dictionary which links their frames to the words and expressions that introduce them, illustrating them with example sentences from the British National Corpus. FrameNet started out for English, but there is already cross-lingual transfer of the framework to German (Erk et al. 2003), Spanish (Subirats/Sato 2004), and Japanese (Ohara et al. 2004). The PropBank project (Palmer/Gildea/Kingsbury 2005) is creating a corpus of text annotated with
425 information about semantic roles by adding a layer of predicate-complement relations to the syntactic structures of the Penn Treebank. In contrast to FrameNet, PropBank defines semantic roles on a per-verb basis, but not across verbs. The PropBank is designed as a broad-coverage resource, covering every instance of a verb in the corpus, to facilitate the development of more general NLP systems.

430 Whole lines of research on semantic roles (partly working on the above databases) have been advanced via the framework of recent and ongoing shared tasks, i.e., competitions where the organisers define a task (and provide the necessary data) in order to compare different approaches to that specific task. In *Senseval* (www.senseval.org) the task is word sense disambiguation. Rich data sets with deep syntactic information are provided for this task, which started out in 1998. Also, within the Conference on Natural Language
435 Learning, the shared task was devoted to semantic role labeling in some of the events.

Diathesis Alternation:

440 Diathesis alternations concern the (systematic) alternative use of frames and semantic roles. Thus, after having assigned semantic information to verb frames, the next natural step is a study of diathesis alternations. Sentences (vi) and (vii) illustrate an example of diathesis alternation, namely the *benefactive alternation* in English, cf. Levin (1993).

(vi) Martha carved a toy for the baby.

(vii) Martha carved the baby a toy.

445 The benefactive alternation is characterised by an alternation between (i) a transitive frame plus a benefactive *for*-PP, and (ii) a double object frame; in addition, the semantic categories of the direct objects in (vi) and (vii) overlap, as well as the semantic categories of the *for*-PP in (vi) and the indirect object in (vii). The alternation is called *systematic*, since it applies to a range of semantically similar verbs, cf. Apresjan (1973). For example, the benefactive alternation transfers to other *build verbs* such as `bake` and `cook`, and
450 *preparation verbs* such as `pour` and `prepare`. This property of regularity makes diathesis alternations an important issue for the creation of verb classes, cf. section 61.2.

Even though a large number of approaches have been concerned with the automatic acquisition of syntactic subcategorisation and there is a substantial amount of work devoted to semantic labeling, few approaches
455 exist for inducing diathesis alternations, and most of those are concerned with case studies. This is probably because an *explicit* definition of diathesis alternations is rarely necessary, while an *implicit* definition (acquiring and applying syntactic combined with semantic subcategorisation) is usually sufficient in the relevant NLP tasks. In the following, three example approaches to the explicit learning of diathesis alternation are presented.

460

McCarthy (2001) introduced to identify which English verbs participate in a diathesis alternation. In a first step, she used the subcategorisation frame acquisition system of Briscoe/Carroll (1997) to extract frequency information on subcategorisation frame types for verbs from the BNC. The subcategorisation frame types were manually linked with the Levin alternations, and thereby defined the verbal alternation candidates.

465 Following the acquisition of the syntactic information, the nominal fillers of the NP and PP complements in the verb-frame tuples were used to define selectional preferences for the respective complement slots. For this step, McCarthy utilised the selectional preference acquisition approach of Minimum Description Length by Li/Abe (1998). In the final step, McCarthy defined two methods to identify the participation of verbs in diathesis alternations: (i) The MDL principle compared the costs of encoding the tree cut models of selectional preferences for the relevant complement slots in the alternation frames. If the cost of combining the models was cheaper than the cost of the separate models, the verb was classified as undergoing the respective alternation. (ii) A similarity-based method calculated the similarity of the two tree cut models with reference to the alternating complement slots for verbs that participated in diathesis alternations. A threshold decided the participation.

470

475 Lapata/Brew (2004) performed a case study on the induction of diathesis alternations, studying the dative and benefactive alternation for English verbs. They used a shallow parser to identify verb frames and their frequencies in the BNC, and defined a simple probabilistic model to generate preferences for the Levin classes.

480 Tsang/Stevenson (2004) based their model of diathesis alternation on distributional similarity between WordNet trees, rather than WordNet classes. The WordNet nominal trees were activated by probability distributions over verb-frame-noun pairs, and standard similarity measures determined the similarity of verb-frame alternations. A threshold defined the participation in an alternation; the work was a case study on the causative alternation.

485 The three above approaches are difficult to compare because they focus on different alternations and are not evaluated on common data. Tsang/Stevenson introduced their approach as an enhancement of McCarthy and showed that their results outperformed the previous approach in the general case (i.e., when applied to random rather than hand-selected data). For empirical linguistics it might be interesting to see further developments of explicit approaches to automatically detect diathesis alternations, especially for other languages than English.

490

61.2 Induction of verb classes from corpora

495 Verb classes categorise verbs into classes such that verbs in the same class are as similar as possible, and verbs in different classes are as dissimilar as possible; the kind of similarity is defined by the creators of the verb classes. For example, syntactic verb classes categorise verbs according to syntactic properties of interest, semantic verb classes categorise verbs according to semantic properties of interest, etc. From a practical point of view, verb classes reduce redundancy in verb descriptions, since they refer to the common properties of the verbs in the classes; in addition, verb classes can predict and refine properties of a verb that received insufficient empirical evidence, by referring to verbs in the same class: under this criterion, a verb classification is especially useful for the pervasive problem of data sparseness in NLP, where little or no knowledge is available for rare events.

500

505 This section is concerned with the *automatic* creation of verb classes, which is supposed to avoid tedious *manual* definitions of the verbs and the classes. The outcome of the creation process depends on several factors, which are summarised as follows:

- the purpose of the classification,
- the choice of the verbs of interest,

510

- the definition of features that describe the verb properties of interest and can be obtained from corpora,
- the choice of an algorithm for class formation and verb assignment to classes, and
- the evaluation of the resulting classification.

In the remainder of this section, we address these parameters. Section 61.2.1 provides an overview of different types of verb classes, section 61.2.2 presents approaches to the automatic creation of verb classes, and section 61.2.3 addresses the evaluation of classifications. As mentioned above, we focus on the empirical acquisition of verb classes and only occasionally refer to manual classifications.

515

61.2.1 Types of verb classes

520

Even though one could think of various linguistic properties to classify verbs, much work on the automatic induction of verb classes has concentrated on verb classes at the syntax-semantics interface. An important reason for this is that few corpora are semantically annotated and provide semantic annotation off-the-shelf (such as FrameNet (Baker/Fillmore/Lowe 1998) and PropBank (Palmer/Gildea/Kingsbury 2005), cf. section 61.1.2). Instead, the automatic construction of syntax-semantics verb classes typically benefits from a long-standing linguistic hypothesis which asserts a tight connection between the lexical meaning of a verb and its behaviour: to a certain extent, the lexical meaning of a verb determines its behaviour, particularly with respect to the choice of its complements, cf. Pinker (1989) and Levin (1993), among others. Even though the meaning-behaviour relationship is not perfect, the following prediction is used: if a verb classification is induced on the basis of features describing verb behaviour, then the resulting behaviour-classification should agree with a semantic classification to a certain extent. From a practical point of view, such verb classes have successfully been applied in NLP. For example, the English verb classification by Levin (1993) was used in NLP applications such as word sense disambiguation (Dorr/Jones 1996), machine translation (Dorr 1997), document classification (Klavans/Kan 1998), and subcategorisation acquisition (Korhonen 2002). In the following, individual approaches to acquire verb classes at the syntax-semantics interface are introduced with respect to their target classification and the choice of features used to empirically model the verb properties of interest.

525

530

535

540

Brent (1991) and Siegel (1998) described approaches to aspectual verb classes, therefore distinguishing between states and events. Both approaches chose features that were indicators of verbal aspect: Brent used syntactic cues such as occurrences of the progressive and adverbial constructions in the verb context; Siegel used a more extensive set of 14 linguistic indicators including Brent's cues, and adding e.g. tense distinctions and prepositional phrases indicating a duration.

545

550

555

560

A major line of approaches to verb classes at the syntax-semantics interface induced empirical information on verb behaviour from corpora, focusing on subcategorisation frames, prepositional phrases, semantic categories of complements, and alternation behaviour, in line with section 61.1. For example, Dorr/Jones (1996) extracted the syntactic patterns from Levin's class descriptions (distinguishing positive and negative instances), and showed that these patterns correspond closely to the affiliation of the verbs with their semantic classes. Merlo/Stevenson (2001) approached three verb classes – unergative, unaccusative, and object-drop verbs – and defined verb features that rely on linguistic heuristics to describe the thematic roles of subjects and objects in transitive and intransitive verb usage. The features included heuristics for transitivity, causativity, animacy, and syntactic features. For example, the degree of animacy of the subject roles was estimated as the ratio of occurrences of pronouns to all subjects for each verb, based on the assumption that unaccusatives occur less frequently with an animate subject when compared to unergative and object-drop verbs. Joanis (2002) and Joanis/Stevenson (2003) presented an extension of their work that approached 14 Levin classes. They defined an extensive feature space including part-of-speech, auxiliary frequency, syntactic categories, and animacy, plus selectional preference features taken from WordNet. Stevenson/Joanis (2003) then applied various approaches to automatic feature selection in order to reduce the feature set to the relevant features, addressing the problem of too many irrelevant features. They reported a semi-supervised chosen set of features based on seed verbs (i.e., representative verbs for the verb classes) as the most reliable choice.

565

Schulte im Walde (2000, 2006) described English/German verbs by probabilities for subcategorisation frames including prepositional phrase types, plus selectional preferences referring to the WordNet/GermaNet top-level synsets. The classification target was semantic verb classes such as 'manner of motion', 'desire', and 'observation'.

570

Esteve Ferrer (2004) acquired verb properties referring to syntactic subcategorisation frames; the target classification referred to the manual Spanish verb classes developed by Vázquez et al. (2000), with the three semantic classes 'trajectory', 'change', and 'attitude', subdivided into 31 subclasses. The Spanish verb classes were similar to Levin's English classes but grouped together different subclasses.

575

Merlo et al. (2002) and Tsang/Stevenson/Merlo (2002) introduced a multi-lingual aspect to the work by Merlo/Stevenson (2001). Merlo et al. (2002) showed that the classification paradigm was applicable to other languages than English by using the same features as defined by Merlo/Stevenson (2001) for the respective classification of Italian verbs. Tsang/Stevenson/Merlo (2002) used the content of Chinese verb features to refine the English verb classification: The English verbs were manually translated into Chinese and given

part-of-speech tags, passive particles, causative particles, and sublexical morphemic properties. Verb tags and particles in Chinese are overt expressions of semantic information that is not expressed as clearly in English. The multi-lingual set of features outperformed either set of monolingual features. The multi-lingual work demonstrates that a) there are features that are useful for the task of verb class acquisition cross-linguistically, and b) an existing feature set in this framework can be extended and improved by exploiting features from a different language.

The overview of the selected approaches illustrates that there are various types of syntax-semantics target classifications, and that the chosen verb features vary accordingly across the targets. On the one hand, a core of features (such as subcategorisation frames) has established itself within the syntax-semantics descriptions; on the other hand, the choice and extraction of empirical features from corpora for verb class creation is still developing.

590

61.2.2 Approaches to acquiring verb classes

Based on the verb descriptions introduced in the previous section, approaches to acquiring verb classes used various supervised or unsupervised methods (cf. article 42) to decide about class membership. For example, Brent (1991) simply defined a confidence interval for his cue frequencies, and a threshold to decide between a stative and an event verb. Siegel (1998), in comparison, applied three supervised machine learning algorithm (logistic regression, decision trees, genetic programming) to his aspectual classification, plus an unsupervised partitioning algorithm which was based on a random assignment and improved by greedy search.

600

Most work in the tradition of Merlo and Stevenson (Merlo/Stevenson 2001; Joanis 2002; Merlo et al. 2002; Tsang/Merlo/Stevenson 2002) used decision trees to establish the verb classes. Schulte im Walde (2000), Stevenson/Joanis (2003) as well as Esteve Ferrer (2004) performed unsupervised clustering, applying agglomerative hierarchical approaches. Schulte im Walde (2006) partitioned verbs into classes by using the unsupervised iterative k-Means algorithm. Even though different classification and clustering approaches were applied to a similar task, it is difficult to compare the above approaches, since none of them was evaluated on common data sets.

605

So far, few approaches addressed the polysemy of verbs by using soft-clustering algorithms and multiple assignment of verbs to classes. For example, Rooth et al. (1999) produced soft semantic clusters for English which at the same time represented a classification on verbs as well as on nouns. The conditioning of the verbs and the nouns on each other was made through hidden classes and the joint probabilities of classes. Verbs and nouns were trained by the Expectation-Maximisation (EM) algorithm. The resulting model defined conditional membership probabilities of each verb and noun in each class.

615

Korhonen/Krymolowski/Marx (2003) used the Information Bottleneck, an iterative soft clustering method based on information-theoretic grounds, to cluster verbs with possible multiple senses. They reported that polysemic verbs with a clear predominant sense or regular polysemy were frequently clustered together. Homonymic verbs or verbs with strong irregular polysemy tended to resist any classification.

620

Last but not least, we find whole projects devoted to the creation of (verb) classes. Prominent examples are – as introduced in section 61.1.2 – WordNet (Fellbaum 1998) which organises English nouns, verbs, adjectives and adverbs into classes of synonyms, and FrameNet (Baker/Fillmore/Lowe 1998) which assigns English verbs, nouns and adjectives to FrameNet frames, referring to common situational knowledge. Even though much of the work in these and other projects is performed manually, selected issues are supported by (semi-)automatic methods.

625

61.2.3 Evaluation of verb classes

630

There is no absolute scheme for automatically evaluating the induced verb classifications. A variety of evaluation measures from diverse areas such as theoretical statistics, machine vision, web-page clustering and coreference resolution do exist, but so far, no generally accepted method has been established. We can distinguish two strands of evaluation methods: (i) methods which address how well the data underlying the

635 verb descriptions are modelled by the resulting classification, and (ii) methods which compare the resulting classification against a gold standards.

The silhouette value (Kaufman/Rousseeuw 1990) represents an example of type (i), evaluating the modelling of the data. It measures which verbs lie well within a class and which verbs are marginal to a class by comparing the verbs' distances to verbs in the same class with the distances to verbs in the neighbour class.
640 The distances between verbs in the same class should be smaller than between verbs in different classes; then the data are well separated by the clustering result. Stevenson/Joanis (2003) and Esteve Ferrer (2004) applied this evaluation as one measure of their classification quality.

645 Evaluation methods of type (i) do not assess whether the clustering result resembles a desired verb classification. In contrast, when applying an evaluation method of type (ii), one needs a gold standard resource of verb classes to compare the clustering result with. Most approaches so far referred to hand-crafted small-scale verb classes which they developed for the purpose of evaluation. Large-scale resources are rare; two instances for English are the Levin classes and WordNet. But even with a gold standard at hand the evaluation task is still difficult, because there are various ways to compare the two sets of classes.
650 Questions such as the following are difficult to answer: how to map the classes within the two sets onto each other, especially when the number of classes is different; whether an evaluation of classes can be reduced to an evaluation of the verb pairs within the classes; how to deal with ambiguity; etc. Schulte im Walde (2003, chapter 4) performed an extensive comparison of various evaluation methods against a gold standard, referring not only to general classification criteria, but also to the task-specific linguistic demands. She
655 determined three evaluation measures as the most appropriate ones to apply: (a) the f-score of a pair-wise precision and recall measure, (b) an adjusted pair-wise precision measure, and (c) the adjusted Rand index. (a) The pair-wise precision and recall measure goes back to a suggestion by Hatzivassiloglou/McKeown (1993) who performed an automatic classification of English adjectives and calculated precision and recall based on common class membership of adjective pairs in the automatic and the gold standard classification.
660 (b) Since the recall value shows strong class size biases, Schulte im Walde and Brew (2002) focused on the precision value and adjusted it by a scaling factor based on the size of the respective verb class. This adjusted pair-wise precision measure (APP) was applied to evaluating verb classes by Schulte im Walde and Brew themselves, and Korhonen/Krymolowski/Marx (2003). (c) The adjusted Rand index (Hubert/Arabie 1985) also measures the agreement between verb pairs in the classes, but is corrected for chance in comparison to the null model that the classes are constituted at random, given the original number of classes and verbs. This
665 measure was applied by Schulte im Walde (2003, 2006), Stevenson/Joanis (2003), and Esteve Ferrer (2004).

The example evaluations illustrate that there is still the need for a generally accepted evaluation method. However, it is also clear that the different approaches on verb class induction have started to agree on a
670 selection of measures.

61.4 Acknowledgements

675 Many thanks to Pia Knöferle, Anna Korhonen, Anke Lüdeling, Alissa Melinger, Sebastian Padó, Nils Reiter, Kristina Spranger, Suzanne Stevenson, and two anonymous reviewers for their feedback on earlier versions of this article.

680 61.5 Literature

Abney, Steven, and Marc Light (1999), Hiding a Semantic Class Hierarchy in a Markov Model. In: *Proceedings of the ACL Workshop on Unsupervised Learning in Natural Language Processing*, 1-8. College Park, MD.

685 Apresjan, Jurij D. (1973), Regular Polysemy. In: *Linguistics* 142, 5-32.

Baker, Collin, Charles Fillmore, and John Lowe (1998), The Berkeley FrameNet Project. In: *Proceedings of the 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics*, 86-90. Montreal, Canada.
690

- 695 Boguraev, Branimir, Ted Briscoe, John Carroll, David Carter, and Claire Grover (1987), The Derivation of a Grammatically-Indexed Lexicon from the Longman Dictionary of Contemporary English. In: *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 193-200. Stanford, CA.
- Brent, Michael R. (1991), Automatic Semantic Classification of Verbs from their Syntactic Contexts: an Implemented Classifier for Stativity. In: *Proceedings of the 5th Conference of the European Chapter of the Association for Computational Linguistics*, 222-226. Berlin, Germany.
- 700 Brent, Michael R. (1993), From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax. In: *Computational Linguistics* 19(2), 243-262.
- Briscoe, Ted, and John Carroll (1993), Generalized Probabilistic LR Parsing for Unification-based Grammars. In: *Computational Linguistics* 19(1), 25-60.
- 705 Briscoe, Ted, and John Carroll (1997), Automatic Extraction of Subcategorization from Corpora. In: *Proceedings of the 5th ACL Conference on Applied Natural Language Processing*, 356-363. Washington, DC.
- 710 Briscoe, Ted, and John Carroll (2002), Robust Accurate Statistical Annotation of General Text. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, 1499-1504. Las Palmas de Gran Canaria, Spain.
- Brockmann, Carsten, and Mirella Lapata (2003), Evaluating and Combining Approaches to Selectional Preference Acquisition. In: *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, 27-34. Budapest, Hungary.
- 715 Carroll, Glenn, and Mats Rooth (1998), Valence Induction with a Head-Lexicalized PCFG. In: *Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing*. Granada, Spain.
- 720 Carroll, John, and Alex Fang (2004), The Automatic Acquisition of Verb Subcategorisations and their Impact on the Performance of an HPSG Parser. In: *Proceedings of the 1st International Joint Conference on Natural Language Processing*, 107-114. Sanya City, China.
- 725 Carroll, John, Guido Minnen, and Ted Briscoe (1998), Can Subcategorisation Probabilities help a Statistical Parser? In: *Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora*, 118-126. Montreal, Canada.
- 730 Chesley, Paula, and Susanne Salmon-Alt (2006), Automatic Extraction of Subcategorization Frames for French. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation*. Genoa, Italy.
- Clark, Steven, and David Weir (2002), Class-based Probability Estimation using a Semantic Hierarchy. In: *Computational Linguistics* 28(2), 187-206.
- 735 Dorr, Bonnie J. (1997), Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation. In: *Machine Translation* 12(4), 271-322.
- 740 Dorr, Bonnie J., and Doug Jones (1996), Role of Word Sense Disambiguation in Lexical Acquisition: Predicting Semantics from Syntactic Cues. In: *Proceedings of the 16th International Conference on Computational Linguistics*, 322-327. Copenhagen, Denmark.
- Dowty, David (1991), Thematic Proto-Roles and Argument Selection. In: *Language* 67, 547-619.
- 745 Eckle-Kohler, Judith (1999), *Linguistic Knowledge for Automatic Lexicon Acquisition from German Text Corpora*. Berlin: Logos Verlag.

- 750 Erk, Katrin, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal (2003), Towards a Resource for Lexical Semantics: A large German Corpus with Extensive Semantic Annotation. In: *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 537-544. Sapporo, Japan.
- 755 Esteve Ferrer, Eva (2004), Towards a Semantic Classification of Spanish Verbs Based on Subcategorisation Information. In: *Proceedings of the Student Research Workshop at the Annual Meeting of the Association for Computational Linguistics*, 37-42. Barcelona, Spain.
- Fellbaum, Christiane (1998), *WordNet – An Electronic Lexical Database*. Cambridge: MIT Press.
- Fillmore, Charles (1982), Frame Semantics. In: *Linguistics in the Morning Calm*, 111-137.
- 760 Fischer, Klaus (1999), Verb Valency – An Attempt at Conceptual Clarification. In: *The Web Journal of Modern Language Linguistics* 4-5. Published by the School of Modern Languages, University of Newcastle upon Tyne.
- 765 Georgala, Effi (2003), A Statistical Grammar Model for Modern Greek: The Context-free Grammar. In: *Proceedings of the 24th Annual Meeting of the Linguistics Department of the Aristotle University of Thessaloniki*. Thessaloniki, Greece.
- Grimshaw, Jane B. (1992), *Argument Structure*. The MIT Press.
- 770 Grishman, Ralph, Catherine Macleod, and Adam Meyers (1994), COMLEX Syntax: Building a Computational Lexicon. In: *Proceedings of the 15th International Conference on Computational Linguistics*, 268-272. Kyoto, Japan.
- 775 Hatzivassiloglou, Vasileios, and Kathleen R. McKeown (1993), Towards the Automatic Identification of Adjectival Scales: Clustering Adjectives According to Meaning. In: *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, 172-182. Columbus, OH.
- 780 Helbig, Gerhard (1992), *Probleme der Valenz- und Kasustheorie*. Number 51 in *Konzepte der Sprach- und Literaturwissenschaft*. Tübingen: Max Niemeyer Verlag.
- Hornby, Albert S. (1985), *Oxford Advanced Learner's Dictionary of Current English*. Oxford University Press.
- 785 Hubert, Lawrence, and Phipps Arabie (1985), Comparing Partitions. In: *Journal of Classification* 2, 193-218.
- Joanis, Eric (2002), *Automatic Verb Classification Using a General Feature Space*. MSc thesis, Department of Computer Science, University of Toronto.
- 790 Joanis, Eric, and Suzanne Stevenson (2003), A General Feature Space for Automatic Verb Classification. In: *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, 163-170. Budapest, Hungary.
- 795 Kaufman, Leonard, and Peter J. Rousseeuw (1990), *Finding Groups in Data - An Introduction to Cluster Analysis*. New York: John Wiley & Sons, Inc.
- Kinyon, Alexandra, and Carlos A. Prolo (2002), Identifying Verb Arguments and their Syntactic Function in the Penn Treebank. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, 1982-1987. Las Palmas de Gran Canaria, Spain.
- 800 Klavans, Judith L., and Min-Yen Kan (1998), The Role of Verbs in Document Analysis. In: *Proceedings of the 17th International Conference on Computational Linguistics*, 680-686. Montreal, Canada.
- 805 Korhonen, Anna (2002), *Subcategorization Acquisition*. PhD thesis, Computer Laboratory, University of Cambridge. Published as Technical Report UCAM-CL-TR-530.

- Korhonen, Anna, Genevieve Gorrell, and Diana McCarthy (2000), Statistical Filtering and Subcategorization Frame Acquisition. In: *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 199-205. Hong Kong, China.
- 810 Korhonen, Anna, Yuval Krymolowski, and Zvika Marx (2003), Clustering Polysemic Subcategorization Frame Distributions Semantically. In: *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 64-71. Sapporo, Japan.
- 815 Lapata, Mirella, and Chris Brew (2004), Verb Class Disambiguation using Informative Priors. In: *Computational Linguistics* 30(1), 45-73.
- Levin, Beth (1993), *English Verb Classes and Alternations*. The University of Chicago Press.
- 820 Li, Hang, and Naoki Abe (1998), Generalizing Case Frames using a Thesaurus and the MDL Principle. In: *Computational Linguistics* 24(2), 217-244.
- 825 de Lima, Erika (2002), *The Automatic Acquisition of Lexical Information from Portuguese Text Corpora with a Probabilistic Context-Free Grammar*. PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Manning, Christopher D. (1993), Automatic Acquisition of a Large Subcategorization Dictionary from Corpora. In: *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, 235-242. Columbus, OH.
- 830 Marcus, Mitchell P., Mary Ann Marcinkiewicz, and Beatrice Santorini (1993), Building a Large Annotated Corpus of English: The Penn Treebank. In: *Computational Linguistics* 19(2), 313-330.
- 835 McCarthy, Diana (2001), *Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Alternations, Subcategorization Frames and Selectional Preferences*. PhD thesis, Department of Informatics, University of Sussex.
- Merlo, Paola, and Suzanne Stevenson (2001), Automatic Verb Classification Based on Statistical Distributions of Argument Structure. In: *Computational Linguistics* 27(3), 373-408.
- 840 Merlo, Paola, Suzanne Stevenson, Vivian Tsang and Gianluca Allaria (2002), A Multilingual Paradigm for Automatic Verb Classification. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 207-214. Philadelphia, PA.
- 845 Meyers, Adam, Catherine Macleod, and Ralph Grishman (1994), Standardization of the Complement Adjunct Distinction. In: *Proceedings of the 7th EURALEX International Congress*. Goteborg, Sweden.
- 850 O'Donovan, Ruth, Michael Burke, Aoife Cahill, Josef van Genabith, and Andy Way (2005), Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II and Penn-III Treebanks. In: *Computational Linguistics* 31(3), 329-365.
- 855 Ohara, Kyoko Hirose, Seiko Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito, and Shun Ishizaki (2004), The Japanese FrameNet Project: An Introduction. In: *Proceedings of the LREC Workshop on 'Building Lexical Resources from Semantically Annotated Corpora'*, 249-254. Lisbon, Portugal.
- Palmer, Martha, Daniel Gildea, and Paul Kingsbury (2005), The Proposition Bank: An Annotated Resource of Semantic Roles. In: *Computational Linguistics* 31(1), 71-106.
- 860 Pinker, Steven (1989), *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge: MIT Press.
- Resnik, Philip (1997), Selectional Preference and Sense Disambiguation. In: *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How*, 52-57. Washington, DC.

- 865 Roland, Douglas, and Daniel Jurafsky (2002), Verb Sense and Verb Subcategorization Probabilities. In: Stevenson, Suzanne, and Paola Merlo (eds), *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*. Amsterdam: John Benjamins, 325-346.
- 870 Rooth, Mats, Stefan Riezler, Detlef Prescher, Glenn Carroll, and Franz Beil (1999), Inducing a Semantically Annotated Lexicon via EM-Based Clustering. In: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 104-111. College Park, MD.
- 875 Sarkar, Anoop, and Daniel Zeman (2000), Automatic Extraction of Subcategorization Frames for Czech. In: *Proceedings of the 18th International Conference on Computational Linguistics*, 691-697. Saarbrücken, Germany.
- Schulte im Walde, Sabine (2000), Clustering Verbs Semantically According to their Alternation Behaviour. In: *Proceedings of the 18th International Conference on Computational Linguistics*, 747-753. Saarbrücken, Germany.
- 880 Schulte im Walde, Sabine (2002a), A Subcategorisation Lexicon for German Verbs induced from a Lexicalised PCFG. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, 1351-1357. Las Palmas de Gran Canaria, Spain.
- 885 Schulte im Walde, Sabine (2002b), Evaluating Verb Subcategorisation Frames learned by a German Statistical Grammar against Manual Definitions in the Duden Dictionary. In: *Proceedings of the 10th EURALEX International Congress*, 187-197. Copenhagen, Denmark.
- 890 Schulte im Walde, Sabine (2003), *Experiments on the Automatic Induction of German Semantic Verb Classes*. PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Schulte im Walde, Sabine (2006), Experiments on the Automatic Induction of German Semantic Verb Classes. In: *Computational Linguistics* 32(2), 159-194.
- 895 Schulte im Walde, Sabine, and Chris Brew (2002), Inducing German Semantic Verb Classes from Purely Syntactic Subcategorisation Information. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 223-230. Philadelphia, PA.
- 900 Siegel, Eric V. (1998). *Linguistic Indicators for Language Understanding: Using Machine Learning Methods to combine Corpus-based Indicators for Aspectual Classification of Clauses*. PhD thesis, Department of Computer Science, Columbia University.
- 905 Spranger, Kristina and Ulrich Heid (2003). A Dutch Chunker as a Basis for the Extraction of Linguistic Knowledge. In: Tanja Gaustad (ed.) *Computational Linguistics in the Netherlands 2002*. Selected Papers from the 13th CLIN Meeting.
- 910 Stevenson, Suzanne, and Eric Joanis (2003), Semi-Supervised Verb Class Discovery Using Noisy Features. In: *Proceedings of the Conference on Computational Natural Language Learning*, 71-78. Edmonton, Alberta.
- 915 Subirats, Carlos, and Hiroaki Sato (2004), Spanish FrameNet and FrameSQL. In: *Proceedings of the LREC Workshop on 'Building Lexical Resources from Semantically Annotated Corpora'*. Lisbon, Portugal.
- Tsang, Vivian, and Suzanne Stevenson (2004), Using Selectional Profile Distance to Detect Verb Alternations. In: *Proceedings of the NAACL Workshop on Computational Lexical Semantics*, 30-37. Boston, MA.
- 920 Tsang, Vivian, Suzanne Stevenson, and Paola Merlo (2002), Cross-linguistic Transfer in Automatic Verb Classification. In: *Proceedings of the 19th International Conference on Computational Linguistics*, 1023-1029. Taipei, Taiwan.

Ushioda, Akira, David A. Evans, Ted Gibson, Alex Waibel (1993), The Automatic Acquisition of Frequencies of Verb Subcategorization Frames from Tagged Corpora. In: *Proceedings of the Workshop on the Acquisition of Lexical Knowledge from Text*, 95-106. Columbus, OH.

925 Vázquez, Gloria, Ana Fernández, Irene Castellón, and María Antonia Martí (2000), *Clasificación Verbal: Alternancias de Diátesis*. Quaderns de Sintagma 3, Universitat de Lleida.

Wauschkuhn, Oliver (1999), *Automatische Extraktion von Verbvalenzen aus deutschen Textkorpora*. PhD thesis, Institut für Informatik, Universität Stuttgart.

930