Resolving polysemy in verbs: Contextualized distributional approach to argument semantics

Anna Rumshisky

Natural language is characterized by a high degree of polysemy, and the majority of content words accept multiple interpretations. Native speakers rely on context to assign the correct sense to each word in an utterance. Natural language processing (NLP) applications, such as the automated word sense disambiguation, require the ability to identify correctly context elements that activate each sense. Our goal in this work is to address the problem of contrasting semantics of the arguments as the source of meaning differentiation for the predicate. We show how the use of insight about the data can help design a targeted distributional approach to this problem. We consider the bidirectional nature of selection processes between the predicate and its arguments and the related problem of overlapping senses. The same sense of a polysemous predicate is often activated by semantically diverse arguments. We introduce the notion of contextualized distributional similarity between such elements, defined with respect to the particular context provided by the predicate. We define the relation of selectional equivalence for predicates, and present an automatic method for clustering both the arguments that activate the same sense of the target predicate and the selectional equivalents of that sense. The proposed method relies exclusively on distributional information, intentionally eschewing the use of human-constructed knowledge sources.

High degree of polysemy in language does not significantly complicate natural language understanding. This is largely due to the fact that within a specific context, each word is usually disambiguated and assigned a single interpretation. The phenomena of lexical ambiguity are exhibited by all the major word classes, and the meaning assigned to the word is determined by a combination of contextual factors relevant for that particular word class. Each element's "meaning potential" (Halliday 1973) is realized and it is locked in the sense it acquires in that context. For example, the meaning assigned to an adjective may be a function of the semantics of the head noun; the meaning of a polysemous noun may be determined by the governing verb or a modifier; and the verbs are typically disambiguated by their dependents and other elements of the syntactic frame. This is illustrated in (1), with the relevant sense given in parentheses:

Rivista di Linguistica 20.1 (2008), pp. 215-240 *(ricevuto nell'ottobre 2008)*

(1) a. *fast* car (*one that is or can be driven quickly*)
b. a *fast* reader (*one who reads quickly*) [1]

c. a rolled-up *newspaper* (*physical object*)
d. a conservative *newspaper* (*organization*)

e. The customer will *absorb* this cost. (*pay*)
f. The customer will *absorb* this information. (*learn*)

In this work, we are concerned with automatic resolution of lexical ambiguity in verbs, especially as it applies to those sense distinctions that can be detected by looking at the semantics of the arguments. While we will mainly discuss polysemous verbs, the same methodology can be applied more generally to any polysemous target word and its 'selectors', i.e. the words with which it forms syntactic dependencies [2]. Our goal is also to show how the use of insight about the data can help design a targeted distributional approach to this problem.

The idea that semantic similarity between words must be reflected in the similarity of habitual contexts in which words occur is fairly obvious and has been formulated in many guises (including the "distributional hypothesis" (Harris 1985), the "strong contextual hypothesis" (Miller & Charles 1991), and even the much-quoted remark from Firth, on knowing the word by the company it keeps (Firth 1957). When applied to the case of lexical ambiguity, it leads one to expect that similar senses of the same word will occur in similar contexts. However, one of the main problems with applying the idea of distributional similarity in computational tasks is that in order to use any kind of generalization based on distributional information, one must be able to identify the sense in which a polysemous word is used in each case.

In this paper, we focus on identifying verbal ambiguities linked directly to the semantics of the words that occur in a particular argument position. As we will see, such words may activate the same sense of the target verb, and yet be quite distinct semantically. In other words, they need to be similar only with respect to the context provided by that verb. We use this intuition to develop a clustering method that relies on contextualized similarity to group such elements together.

The rest of the paper is organized as follows. We first outline the problem of sense detection for verbs, and review some of the relevant ideas from the techniques used in manual construction of knowledge sources. We then review briefly some of the distributional approaches to sense disambiguation and discuss some problems with the notion of word sense, insofar as it relates to selection processes and sense assignment. We then reformulate the problem of measuring semantic similarity with respect to a particular context and outline a distributional method for identifying semantically diverse arguments that activate the same sense of a polysemous word. We demonstrate the way the proposed solution works by examining more closely some intermediate results.

1. Resolving Polysemy in Context

Within the scope of a sentence, the meaning that gets assigned to a word is usually determined by a combination of two factors: (1) syntactic frame into which the word is embedded, and (2) semantics of the words with which it forms syntactic dependencies. We will use the term 'selector' to refer to such words, regardless of whether the target word is the headword or the dependent in the syntactic relation. Syntactic frame should be understood broadly as extending to minor categories (such as adverbials, locatives, temporal adjuncts, etc.) and subphrasal cues (genitives, partitives, negatives, bare plural/determiner distinction, infinitivals, etc.). The set of all 'usage contexts' in which a polysemous word occurs can usually be split into groups where each group roughly corresponds to a distinct 'sense'. In some cases, a more extended context is required to resolve the indeterminacy. But typically, a clause or a sentence context is sufficient for disambiguation.

To illustrate the contribution of different context parameters to disambiguation, consider the verbs in (2) and (3). Syntactic patterns for the verb *deny* in (2) are usually sufficient to disambiguate between the two dominant senses: (i) 'refuse to grant' and (ii) 'proclaim false'[3].

(2) <u>Syntactic frame:</u>
a. The authorities *denied* that there is an alternative. [that-CLAUSE]
The authorities *denied* these charges. [NP]
('proclaim false')
b. The authorities *denied* the Prime Minister the visa. [NP] [NP]
The authorities *denied* the visa to the Prime Minister. [NP] [to-PP]
('refuse to grant')

For the senses of *fire*, *absorb*, *treat* and *explain* shown in (3), contrasting argument and/or adjunct semantics is the sole source of meaning differentiation. The relevant argument type is shown in brackets and the corresponding sense in parentheses:

(3) <u>Semantics of the arguments and adjuncts/adverbials:</u>

a. The general *fired* four *lieutenant-colonels*. [PERSON] ('dismiss')
The general *fired* four *rounds*. [PHYSOBJ] ('shoot')

b. The customer will *absorb* this *cost*. [ASSET] ('pay')
The customer will *absorb* this *information*. [INFORMATION] ('learn')

c. Peter *treated* Mary with *antibiotics*. [with MEDICATION] ('medical')
Peter *treated* Mary with *respect*. [with QUALITY] ('human relations')

d. This new *booklet explains* our strategy. [INFORMATION] ('describe, clarify')
This new *development explains* our strategy. [EVENT] ('be the reason for')

Establishing a set of senses available to a particular lexical item and (to some extent) specifying which context elements typically activate each sense forms the basis of any lexicographic endeavour. Several current resource-oriented projects undertake to formalize this procedure, utilizing different context specifications. FrameNet (Ruppenhofer et al. 2006) attempts to organize lexical information in terms of script-like semantic frames, with semantic and syntactic combinatorial possibilities specified for each frame-evoking 'lexical unit' (word/sense pairing). Different senses of a polysemous word are associated with different frames. FrameNet uses Fillmore's case roles to represent semantics of the arguments. Case roles ('frame elements') are derived on ad-hoc basis for each frame. Context specification for each lexical unit contains such case roles (e.g. Avenger, Punishment, Offender, Injury, etc. for the Revenge frame) and their syntactic realizations, including grammatical function (Object, Dependent, External Argument (= Subject)), etc.), and phrase type (e.g. NP, PP, PPto, VPfin, VPing, VPto, etc.). Core frame elements represent semantic requirements of the target lexical unit, some of which may not be actually expressed in the sentence.

Corpus Pattern Analysis (CPA) (Hanks & Pustejovsky 2005) attempts to catalog prototypical norms of usage for individual words, specifying them in terms of context patterns. Each pattern gives a combination of surface textual clues and argument specifications. CPA uses the extended notion of syntactic frame, as outlined above. Semantics of the arguments is represented either through a set of shallow semantic types representing basic semantic features (e.g. Person, Location, PhysObj, Abstract, Event, etc.) or extensionally through 'lexical sets', which are effectively collections of lexical items.

For example, for the verb *absorb* in (3b), the following pattern specification would be recorded [4]:

(4) a. [[Person]] absorb [[LEXSET Asset: tax, cost,...]]
b. [[Person]] absorb [[Information]]

As a corpus analysis technique, CPA derives from the analysis of large corpora for lexicographic purposes, of the kind that was used for compiling the Cobuild dictionary (Sinclair & Hanks 1987). For each target word, a lexicographer groups similar contexts of occurrence together and gives a pattern specification for each group. Several context patterns may represent a single sense, with patterns varying in syntactic structure and/or the encoding of semantic roles relative to the described event.

While manually constructed knowledge sources such as machine readable dictionaries or thesauri are extremely valuable for some tasks, they simply do not endeavor to specify the context parameters relevant for sense distinction. At the same time, the resources that do attempt it are often incomplete. For example, FrameNet, which proceeds with sense analysis frame by frame, often specifies only one out of several senses for each lexical item [5]. The CPA approach, which relies on full context analysis for each word, is painstakingly slow and consequently lacks coverage. On the other hand, in WordNet (Fellbaum 1998), for example, despite its wide coverage, sense distinctions are frequently not sufficiently founded in actual patterns of use and no real attempt is made to specify the relevant context parameters. Also, as we will see below, the requisite semantic information is very context-dependent and difficult to capture during lexicographic analysis.

These factors make it desirable to have the ability to detect the relevant context parameters relying exclusively on the distributional information. However, we think it beneficial to separate the two factors that contribute to disambiguation. Identifying the first factor, namely, syntactic frame elements relevant for predicate sense assignment, is linked directly to the success of parsing. If a reliable parse is obtained for all occurrences of the target predicate, each instance can then be assigned to the appropriate sense group. Detecting the second factor, that is, the relevant components in the semantics of the arguments, is a much less straightforward task. In the next section, we briefly review some of the approaches to resolving polysemy based on overall distributional similarity. We then examine the considerations that come into play when the ambiguity of the predicate is resolved based solely on the semantics of the arguments.

2. Distributional Similarity in NLP

The notion of distributional similarity is used in NLP in a number of tasks, including areas such as word sense disambiguation (WSD), sense induction, automatic thesaurus construction, selectional preference acquisition, and semantic role labeling. It is used to identify semantically similar words (as in thesaurus construction) or similar uses of the same word (as in WSD and sense induction).

Resulting clusters of distributionally similar words are often seen as a means to address the problem of data sparsity faced by many NLP tasks. The problem is that a lot of fairly common content words occur very infrequently in actual texts. Their counts thus can not be used to reliably predict their behavior, which is especially problematic since a significant percentage of actual texts is made up of precisely such rare events. Dunning (1993) reports, for example, that words with frequency of less than one in 50,000 make up 20-30% of news-wire reports. With respect to word co-occurrence, the problem is exacerbated further, since the number of possible joint events is much larger than the number of events actually encountered in texts. Generalizing across clusters allows one to model rare events, thereby alleviating the problems caused by sparsity in “middle layer” NLP tasks, including, for example, any number of parsing-related problems, such as resolving PP-attachment, scope of modification, nominal compounds, and other kinds of structural ambiguity.

One of the main challenges in using distributional similarity to generalize over word classes is that for a polysemous word, generalizations must apply to different ‘senses’, rather than to all of its occurrences uniformly. In the absence of a semantically tagged corpus, obtaining frequency counts for each sense in a straightforward manner is impossible. Consequently, semantics of the arguments is often represented using information derived from external knowledge sources, such as FrameNet, machine-readable dictionaries, WordNet, etc. (cf. Navigli et al. 2007, Pradhan et al. 2007, Mihalcea et al. 2004, Agirre & Martinez 2004).

To illustrate this problem, consider acquiring selectional preferences for a given verb from a corpus. In order to do that, one needs to obtain counts for different semantic categories of nouns with which that verb occurs. That is impossible, unless all polysemous nouns occurring with that verb have been properly disambiguated. Since sense-tagged data is very costly to produce, raw text must be used in many such tasks. As a result, one often has to settle for fairly imperfect solutions. For example, Resnik (1996) computed selectional preferences for verbs by normalizing the count for each noun by the number of senses it has in WordNet.

In sense induction and automatic thesaurus construction literature, the goal is to create an alternative to external knowledge sources. Therefore, clusters of similar words are usually obtained solely on the basis of distributional information (Grefenstette 1994, Schütze 1998, Pantel & Lin 2002, Dorow & Widdows 2003, Velldal 2005). Each word's representation is linked to a set of contexts in which it occurs in a corpus. Context is typically represented as a feature vector, where each feature corresponds to some context element. The value of each feature is the frequency with which that element is encountered together with the target word. A word may be represented as a feature vector combining all the context features or as a probability distribution on the joint events of occurrence of the target word with each context element. Alternatively, all words (including the target word) may be regarded as nodes in a co-occurrence graph, where the co-occurring context elements are represented by the neighboring nodes, and the frequency of the co-occurrence is the weight assigned to the corresponding edge (Widdows & Dorow 2002, Agirre et al. 2006). Some approaches use distributional features based on bag-of-words style co-occurrence statistics (Schütze 1998, Gale et al. 1993, Widdows & Dorow 2002), others use context representations that incorporate syntactic information, and sometimes semantic information from external sources (Grefenstette 1994, Lin 1998, Pantel & Lin 2002). In the latter case, each distributional feature may correspond to a grammatical relation populated with a particular word or an entity type.

Solving the problem of polysemy amounts to separating out the occurrences corresponding to each sense from the distributional representation of the target word. Typically, this problem is resolved by either (1) clustering similar occurrence contexts for each word, or (2) clustering the actual words whose overall distributional profiles are similar (Schuetze 1998, Grefenstette 1994, Lin 1998)[6]. Top-K words whose overall distributional profiles are most similar to the target are often grouped into tight clusters that represent the target's senses (Grefenstette 1994, Lin 1998, Widdows & Dorow 2002). Alternatively, each word may be associated with multiple clusters, as in soft clustering (Pantel and Lin 2002, Velldal 2005). Pantel & Lin (2002), for example, suggest removing from a distributional representation of a word the features associated with the closest cluster of mostly monosemous words and then assigning the word to the next closest cluster based on the features that remain. Some recent works have also attempted to bring attention to the contextualized nature of semantic similarity between words and utilize this idea in designing computational approaches to polysemy resolution (Allegrini et al. 2003, Gamallo et al. 2005, Rumshisky et al. 2007).

3. Selection and Sense Assignment

Computational approaches to word sense disambiguation typically assume that each word in an utterance is assigned a sense from an inventory of senses. This is clearly a simplification of what actually happens when the meaning of a complex expression is computed. Consider a polysemous target predicate with certain semantic preferences. In a given argument position, different senses of that predicate will select for different semantic features. Thus, in (3b), the 'pay' sense of *absorb* selects for Asset in direct object position, while the 'learn' sense selects for Information. Similarly, the 'dismiss' sense of *fire* in (3a) selects for Person, while the 'shoot' sense selects for PhysObj, [+projectile].

Selection is effectively a 'bidirectional' process through which a particular interpretation is assigned both to the predicate and to its arguments. For example, in (5), the noun *rounds* is ambiguous between the TimePeriod and PhysObj, and it is disambiguated by the 'shoot' sense of *fire*, while simultaneously activating that sense for the predicate.

(5) The general *fired* four *rounds*.

The predicate and its argument in such cases essentially form a 'minimal disambiguation unit' which does not require any additional context for all the elements within the unit to be disambiguated.

3.1. Sense-Activating Argument Sets

The same sense of the predicate may be activated by a number of semantically diverse arguments. For some of them, the relevant semantic feature will be central to their meaning. For others, it will be merely a contextual interpretation that they permit. Effectively, each sense of the target predicate may be seen to induce an ad-hoc semantic category in the relevant argument position.

For example, consider two of the senses of the phrasal verb *take on*: (i) 'tackle an adversary' and (ii) 'acquire a quality'. Some of the lexical items that occur in direct object position for these two senses are given in (6).

(6) a. *Tackle an adversary*:
competition, rival, enemy, opponent, team, government, world.
b. *Acquire a quality*:
shape, meaning, color, form, dimension, reality, significance, identity, appearance, characteristic, flavor.

The nouns in each argument set are semantically quite distinct, and yet they activate the same sense of the predicate. The context provided by the predicate merely selects a particular aspect of their sense. As often happens, the argument sets consist of a number of 'core' elements for which it is a central component of their meaning and some 'satellite' members for which the requisite component may be peripheral. Thus, in the first argument set, the [+adversary] component is central for *enemy*, *rival*, *opponent* and *competition*, while *government* and *world* merely allow this interpretation due to animacy/agency.

Core members of the argument set may be polysemous and require the 'bidirectional selection' process in order to activate the appropriate sense of the predicate. But notice that the interpretive work that is done in (7a) and (7b), for example, is quite different.

(7) a. Are you willing to *take on* the competition?
b. Are you willing to *take on* the government?

While both words activate the same sense of *take on*, *competition* will merely be disambiguated between the EVENT reading and the ANIMATE, [+adversary] reading. For *government*, the [+adversary] reading will be coercively imposed by the predicate and is effectively accidental.

Another observation to make is that different aspects of meaning may be relevant for different dependencies the word enters into. For example, consider the use of the noun *opponent* with the verbs *take on* and *know* in (8a).

(8) a. It is much harder to *take on* the opponent you *know* personally.
b. It is much harder to *take on* the student you *know* personally.

FrameNet gives two senses for the verb *know*: (1) the 'familiarity' sense (this is the sense in which you know people and places) and (2) the 'awareness' sense (this is the sense in which you know propositional content). In (8a), the word *opponent* activates the familiarity reading for *know* and the adversary reading for *take on*. While the second operation requires the [+adversary] component, the PERSON reading is sufficient for the first operation. This difference is made more apparent by the fact that in (8b), for example, the word *student* which is lacking the [+adversary] component, activates a different sense of *take on*. Effectively, the relevant semantic component in the interpretation of *opponent* changes according to the context provided by the verb.

3.2. Selector-Based Sense Separation

In case of homonymy, semantic components selected for by different senses may be sufficiently distinct. In that case, overall distributional similarity between arguments may be sufficient to group together the relevant lexical items. For example, *file* in the sense of 'smooth' (e.g. *file nails, edges*, etc.) is easily distinguished from the cluster of senses related to *filing papers*. But in case of polysemy, separating different senses of the verb is notoriously hard even for a trained human eye.

This problem has been the subject of extensive study in lexical semantics aiming to address questions such as when the context selects a distinct sense and when it merely modulates the meaning, what is the regular relationship between related senses, and what compositional processes are involved in sense selection (Pustejovsky 1995, Cruse 1995, Apresjan 1973). These considerations have also been the concern of the computational community working on sense disambiguation, where evaluation requires having a uniformly accepted sense inventory. At recent Senseval efforts (Mihalcea et al. 2004, Snyder & Palmer 2004, Preiss & Yarowsky 2001), the choice of a sense inventory frequently presented problems. One of the proposed views has been that it is impossible to establish a standard inventory of senses independent of the task for which they are used (cf. Agirre & Edmonds 2006, Kilgarriff 1997). Attempts have been made to create coarse-grained sense inventories (Navigli 2006, Hovy et al. 2006, Palmer et al. 2007). Inventories derived from WordNet by using small-scale corpus analysis and by automatic mapping to top entries in Oxford Dictionary of English were used in the most recent workshop on semantic evaluation, Semeval-2007 (Agirre et al. 2007). In lexicography, "lumping and splitting" senses during dictionary construction – i.e. deciding when to describe a set of usages as a separate sense – is also a well-known problem (Hanks & Pustejovsky

2005, Kilgarriff 1997). It is often resolved on ad-hoc basis, resulting in numerous cases of "overlapping senses", i.e. instances when the same occurrence may fall under more than one sense category simultaneously. Any analysis of verb polysemy runs into this problem especially.

Consider what happens if we need to determine which selectors are likely to activate what sense, keeping in mind that at least some of the verb's senses will be interrelated. Typically, corpus occurrences of a polysemous verb cluster into 2-10 groups, each roughly corresponding to a sense[7]. For each of these groups, one usually finds a lot of cases where the sense distinctions are clear-cut and easily discernable. However, whenever two senses are related, there are usually some boundary cases when it is not clear which sense of the predicate is used. Thus, in a given argument position, three kinds of selectors are possible:

(i) GOOD DISAMBIGUATORS: selectors that immediately pick one sense of the target. These can be monosemous or polysemous themselves. When such selector is polysemous, its other sense(s) just never occur with the other sense of the target verb. Disambiguation is achieved through bidirectional selection, as in "fire four rounds" in (5).

(ii) POOR DISAMBIGUATORS: selectors that may be used with either sense and require more context to be disambiguated themselves (bidirectional selection doesn't work). For example, *assuming a position* may equally likely mean 'taking on a post, adopting a particular bodily posture', 'occupying a certain point in space', or 'presupposing a certain mental attitude', etc.

(iii) BOUNDARY CASES: the choice between two senses of the target is in fact impossible to make (i.e. selector activates both senses at once).

For example, for the subject position with the verb *show* in (9), *survey* and *photo* are good disambiguators, while *graph* is a clear example of a boundary case.

(9) a. The photo *shows* Sir John flanked by Lt Lampard. ('pictorially represent')
b. The survey *shows* signs of improvement in the second quarter. ('demonstrate by evidence or argument')
c. The graph *showed* an overall decrease in weight. (both senses?)

Boundary cases are obviously identified as such only when each individual sense can be clearly defined, that is, when good disambiguators for each sense are very common. For that reason, such cases are better construed as instances of 'multiple selection' (i.e. simultaneous activation of both senses), and not merely as evidence for overlapping sense definitions. Interestingly, even syntactic pattern can not always overrule the interpretation intrinsic to some selectors. For example, in (10), it is virtually impossible to resolve *deny* between 'refuse to grant' and 'proclaim false':

(10) a. Elders are often *denied* the status of adulthood
b. Philosophers have *denied* the autonomy to women

In (11), on the other hand, the selector itself is polysemous, with two interpretations available for it, and it needs to be disambiguated by context before it can activate the appropriate sense of the predicate.

(11) a. *deny* the traditional *view* ('proclaim false')
b. *deny* the *view* of the ocean ('refuse to grant')

In the following sections, we discuss how these considerations can be taken into account when designing a computational strategy for automatic sense detection.

4. Contextualized Similarity

The goal of a similarity measure is to allow us to tell automatically whether one word is "like" the other. But whether one word is like the other may vary, depending on the particular task. If our task is to determine the meaning of a predicate by looking at its arguments, two words in the same argument position will be "like" each other only if they pick the same sense of the predicate. We can capture this intuition by defining a measure aimed to assess 'contextualized similarity', i.e. similarity between two lexical items with respect to a particular context.

We adopt a context representation based on the notion of grammatical relation as it is used in distributional similarity literature (see, e.g. Lin 1998, Hindle 1990). A grammatical relation is a tuple (w_1, R, w_2), where R denotes the type of grammatical dependency between the words w_1 and w_2. A context is a set of such tuples, as extracted from a single instance of occurrence of the target word. For example:

(12) sentence: Their life *took on a different meaning*.
context(*meaning*): {(*take on*, obj, *meaning*), (*meaning*, modifier, *different*)}

In the following discussion, we will use the term 'context' to refer to a singleton, i.e. a single populated syntactic relation. For example, the verb *take on* and the relation of direct object above define a particular context of occurrence for the noun *meaning*. We will use an abbreviated notation for such singleton contexts: (*take on*, obj), (*different*, modifier^{-1}), and so on.

At its most basic, distributional similarity between frequency profiles of two words should reflect to what extent the contexts in which the two words occur overlap. Similarity between two words may be expressed as the frequency of their occurrence in identical contexts, relative to the average of their overall frequencies. Since the two words may have very different corpus frequencies, some normalization is also typically used. The result is a function of relation tuple frequency, typically referred to as the 'weighting' or the 'association score' between the word and the context attribute [8].

Defined in this manner, distributional similarity will be high for lexical items whose overall distributional profiles are similar. This will be the case for words which are semantically very close in their dominant, most frequent sense. Or, in a less likely case, it may be that most of their senses are similar, and have similar relative frequencies. When several nouns from a given argument set activate the same sense of a polysemous verb, high similarity values may be obtained for the elements of the semantically uniform core of this argument set (if such a core is present). On the other hand, polysemous core elements for which the relevant semantic component is not dominant, as well as peripheral elements of this argument set, will slip through the cracks.

Hindle (1990) remarks that while one can *have* and *sell* both *beer* and *wine*, it's the fact that you can *drink* both of them that makes them semantically close. In other words, when computing semantic similarity based on distributional behavior, some contexts are, to quote Orwell, "more equal than others". The reason we know that two words are used similarly in a given context is that there is a number of other contexts in language where they are used in the same way. Such 'licensing contexts' license the use of these lexical items with the same sense of the target word.

Consider, for example, selectors of the two senses of *take on* in (6). Table 1 shows some of the contexts in which these selectors occur [9]. The fact that both *significance* and *shape* occur as direct objects of such verbs as *retain*, *obscure*, and *acquire* allows them to activate the 'acquire a quality' interpretation for *take on*. Note that licensing contexts do not need to be syntactically parallel to the target context. So (*struggle*, pp_against) may select for the same semantic property as the *tackle an adversary* sense of (*take on*, obj).

When computing contextualized similarity for two selectors, we would like to give higher weights to the terms that correspond to the licensing contexts. Consider, for example, using the contexts shown in Table 1 to compute similarity between *competition* and *government* as direct objects of *take on* [10]. Their association scores with contexts similar to the target context must have a higher weight than their association scores with non-similar contexts, i.e. (*threaten*, obj), (*confront*, obj) and (*struggle*, pp_against) should carry a higher weight than (*prize*, n_modifier) or (*the*, det). When both selectors occur in an unrelated context, the latter may in fact activate a completely different reading for each of them. For example, in the phrase "competition prize" *competition* is interpreted as an Event, and not as Animate, [+adversary]. Consequently, the fact that both *government* and *competition* occur as nominal modifiers of *prize* should not be regarded as evidence of their similarity as direct objects of *take on*.

Table 1. Sample licensing contexts for selectors of *take on*.

target context: (take on, obj)					
phrase	context	selectors, P(selector | context)			
		significance	shape	competition	government
retain __	(retain, obj)	.0030	.0030	.0000	.0000
obscure __	(obscure, obj)	.0016	.0043	.0000	.0000
acquire __	(acquire, obj)	.0043	.0006	.0000	.0000
threaten __	(threaten, obj)	.0000	.0008	.0008	.0057
confront __	(confront, obj)	.0000	.0000	.0009	.0104
struggle against __	(struggle, pp_against)	.0000	.0000	.0008	.0089
__ prize	(prize, n_modifier)	.0000	.0000	.0069	.0005
the __	(the, det)	.0000	.0000	.0000	.0000

Computing similarity between contexts thus poses a separate problem. It is clearly incorrect to use overall distributional similarity between context-defining words to determine how close two contexts are. In order to be considered similar, two contexts must be similar with respect to their selectional properties, i.e. select for the same semantic component in the specified argument position. We introduce the notion of 'selectional equivalence' below as a way of addressing this problem.

5 Selectional Equivalence

Selectional equivalence is defined for two verbs with respect to a particular argument position and a particular sense for each verb. If nouns can be organized into lexical sets sharing a semantic feature, verbs can be organized into selectional equivalence sets, with arguments sharing a semantic feature.

A lexical item w_1 is a 'selectional equivalent' ('contextual synonym') of lexical item w_2 with respect to a certain grammatical relation R if one of its senses selects for the same aspect of meaning as one of the senses of w_2 in the argument position defined by R. Selectional equivalents do not need to be synonyms or antonyms of each other. Their equivalence is only in terms of the aspect of meaning they select. Verbs that are selectionally equivalent to one of the senses of the target verb effectively form a subset of all licensing contexts for that sense.

If we can measure how close two contexts are with respect to the target context, selectional equivalents can be grouped into clusters representing different senses of the target verb. Resulting clusters can then be used to determine how likely each selector is to be associated with that sense. We outline this procedure below in Section 6. Clusters of selectional equivalents obtained for selected senses of *take on*, *launch*, and *deny* are shown in (13).

(13) a. *take on* ('acquire a quality')
acquire, obscure, assume, retain, possess
b. *launch* ('begin')
organize, mastermind, spearhead, orchestrate, mount, commence, initiate, instigate, intensify, complete, undertake
c. *deny* ('proclaim false')
confirm, disclose, conceal, reveal, uncover, corroborate, rebut, substantiate, disprove, refute, contradict, retract, furnish, gather, cite, collate, produce, detail, present, summarize, suppress, publicize
d. *deny* ('refuse to grant')
refuse, grant, revoke, obtain, withhold

'Selectional equivalence' thus implies a specific kind of semantic similarity, which overlaps only partially with what manually constructed resources typically aim to capture. In FrameNet, for example, selectionally equivalent verbs may belong to the same frame, or to the frames related through some frame-to-frame relation, such as frame inheritance or the Using relation. This is reasonable, since one would expect semantically uniform core elements to be similar when the verbs that operate on them are from the same situational frame. For example, *deny* and *confirm* in (13c) both evoke the same *Statement* frame; *disclose* and *reveal* evoke the frame which inherits from *Statement*. On the other hand, pairs such as *obscure* and *assume* in (13a) are not likely to evoke related frames. The same partial overlap can be observed with Levin classes and WordNet categories.

In order to obtain clusters of selectional equivalents for each sense of the target verb, we need to be able to measure to what extent two verb senses share selectional properties. This measure of selectional equivalence effectively mirrors contextualized similarity as defined for selectors. The idea is to take all selectors that occur in the specified argument position with the target verb, identify the verbs that occur with these selectors, and cluster them according to the sense of the target with which they share selectional properties. Our model involves the assumption that two verbs tend to be selectionally close with respect to just one of their senses. Similarity between two verbs is estimated based on selectors that, for each of them, consistently activate the sense which is selectionally equivalent to one of the target's senses.

In the next section, we outline the overall architecture of the algorithm and discuss in more detail the choice of reliable selectors. We then look at some results of the similarity computation based on the obtained selector lists.

6. Algorithm Architecture

Consider a bipartite graph where one set of vertices corresponds to headwords and the other to dependents, under a relation R. Each relation can be viewed as a function mapping from headwords to dependents [11]. The relation is defined by a set of tuples (w, R, w'), where w is the head, and w' is the dependent. The inverse of each relation is then a set of tuples (w', R^{-1}, w). Given a corpus, we proceed as follows:

1. Identify the set of selectors with which the target verb occurs in relation R, and then take the inverse image of that set under

the relation R^{-1}. For example, for the target verb $t = acquire$, R = obj, the first operation gives the set of nouns that occur in direct object position with *acquire*. The second operation gives us the set of potential selectional equivalents for different senses of *acquire*. We discard the verbs that occur with a given selector only once in the corpus, under the assumption that such occurrences are spurious.

2. For every word in the set of candidates for selectional equivalency, we obtain a set of reliable selectors (i.e. selectors that pick the same interpretation both for the target and for potential selectional equivalent). The resulting short list of selectors chosen with respect to the target context is then used to construct a contextualized vector representation for each potential selectional equivalent. The choice of reliable selectors is discussed in Section 6.1.
3. Compute similarity between each pair of potential selectional equivalents using the obtained contextualized vector representations (see Section 6.2 below for details).
4. Perform clustering to produce clusters of selectional equivalents for each sense of the target verb [12].
5. Using the obtained clusters, estimate which sense each of the target's selectors is likely to occur with.

For each of the target's selectors s in grammatical relation R, we can compute an association score for each of the chosen clusters C:

$$assoc(s, C) = \sum_{w \in C} P(s|wR)$$

The resulting score indicates how likely selector s is to pick the sense of the target associated with C. The difference between the scores obtained for different senses with a given selector indicates how strongly that selector tends to prefer one of the senses. If the difference is small, the selector may be equally likely to select for either of the senses, or it may select for both senses at once.

6.1. Identifying reliable selectors

Reliable selectors are good disambiguators for both verbs, and they select the same interpretation both for the target and for the potential selectional equivalent. They are reliable in this behavior, i.e. they are not likely to occur with the other senses of each verb.

If two verbs are selectionally equivalent with respect to one of their senses, and a selector occurs fairly frequently with each verb, several explanations are possible. Consider the verbs *take on*, *acquire*, and *possess*, all of which have a sense that selects for QUALITY in direct object position:

(i) A selector could reliably activate the appropriate sense for each verb:
   a. *take on/acquire* a new *importance*
(ii) (*Parallel Sense Distinctions*.) If the verbs have more than one selectionally equivalent sense, a selector could activate the wrong pair of senses:
   a. *acquire/possess* a new *significance* (QUALITY)
   b. *acquire/possess* a powerful *weapon* (POSSESSION)
   The word *weapon* above activates the *ownership* sense for both verbs, rather than the sense involving *having a quality*.
(iii) (*Selector Polysemy*.) Different senses of that selector may activate unrelated interpretations for the two verbs:
   a. *take on* a greater *share* of the load
   b. *acquire* the *shares* of the company

In our model, we make the assumption that the first case is the dominant one, while the other two cases are much more rare. Under such conditions, selectors that occur "frequently enough" with both verbs must be the ones that pick the corresponding sense for each verb. Frequency distribution on the verb senses also remains important, since the relevant sense may be much more prominent for one verb than for the other.

For every word in the set of candidates for selectional equivalence, a set of reliable selectors can be obtained as follows:

1. Take all selectors which occur both with the target and the selectional equivalent. Compute two conditional probability scores for each selector s: $P(s \mid wR)$ and $P(s \mid tR)$, where w is the potential selectional equivalent, t is the target verb, and R is the grammatical relation. For example, $P(\mathit{importance} \mid \mathit{acquire}, \text{obj})$ and $P(\mathit{importance} \mid \mathit{take\ on}, \text{obj})$.

Table 2. Top-15 selectors chosen for *take on* and *acquire* (left) and *assume* and *reveal* (right). Good disambiguators for the target semantic component QUALITY are italicized.

(take_on, obj) vs. (acquire, obj)			(assume, obj) vs. (reveal, obj)		
Selector	$P(s \mid v_1R_1)$	$P(s \mid v_2R_2)$	Selector	$P(s \mid v_1R_1)$	$P(s \mid v_2R_2)$
meaning	0.0123	0.0068	*importance*	0.0368	0.0020
significance	0.0114	0.0048	*role*	0.0473	0.0010
responsibility	0.0211	0.0024	*power*	0.0178	0.0020
work	0.0077	0.0046	*position*	0.0124	0.0026
share	0.0015	0.0206	existence	0.0066	0.0042
role	0.0340	0.0007	*name*	0.0033	0.0065
form	0.0071	0.0032	number	0.0018	0.0109
land	0.0015	0.0141	level	0.0040	0.0048
power	0.0017	0.0121	*significance*	0.0153	0.0010
character	0.0054	0.0032	knowledge	0.0070	0.0022
status	0.0017	0.0092	*character*	0.0055	0.0026
skill	0.0004	0.0351	*identity*	0.0018	0.0071
importance	0.0047	0.0032	*form*	0.0197	0.0006
dimension	0.0062	0.0022	*proportion*	0.0109	0.0010
business	0.0015	0.0073	rate	0.0084	0.0012

2. The reliable selectors will have relatively high conditional probabilities with both words. Conditional probability value will depend on how frequent the appropriate sense is for each of the two words. We use the geometric mean of the above conditional probabilities, and choose the top-K selectors that maximize it.

Table 2 shows selectors chosen for the direct object position of two verb pairs, *take on* and *acquire*, and *assume* and *reveal* [13]. Selector quality for each pair of contexts is estimated as the geometric mean of conditional probabilities. This induces a sorting order with the sequence of equivalence classes located along the hyperbolic curves. Figure 1 illustrates how selectors are picked, with conditional probabilities for the target $P(s \mid tR)$ along the x-axis, and conditional probabilities for the selectional equivalent $P(s \mid wR)$ along the y-axis. Selectors that were manually identified as good disambiguators are depicted in gray.

Clearly, automatically identifying all good disambiguators is not feasible. Our goal is to choose enough selectors correctly so that selectional equivalents of the same sense can be grouped together. For the top-15 selectors chosen by this method, Table 2 shows good disambiguators in italic; higher geometric means are obtained for the top correct choices.

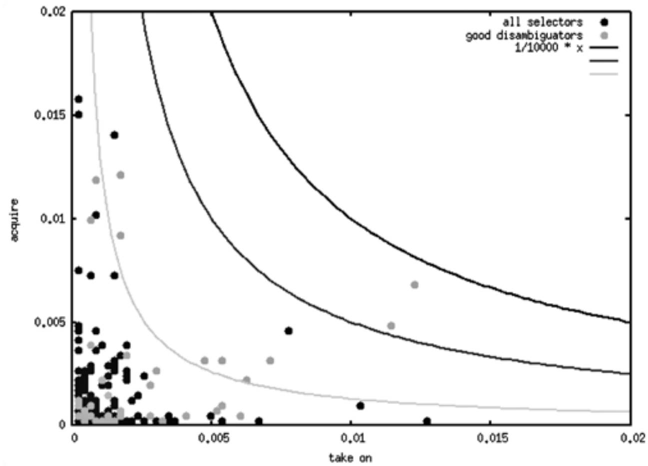

Figure 1. Choosing selectors for the verb pair *take on/acquire*.

6.2. Similarity computation

We compute contextualized similarity for two potential selectional equivalents w_1 and w_2 as the sum of minima [14] of conditional probabilities for every reliable selector in the list obtained for w_1 and w_2:

(1) $$csim(w_1, w_2) = \sum_{s \in S_1 \cap S_2} \min\left(P(s|w_1R), P(s|w_2R)\right)$$

where t is the target verb, R is the grammatical relation, and S_i is a set of top-K selectors that pick the same sense for w_i and t. S_i is approximated by the list of selectors obtained for w_i as described in Section 6.1.

Unlike the standard numerical extensions of Jaccard and Dice measures, we do not normalize the sum of minima either by the size of the union, or by the average size of each set. We do so to avoid obtaining high similarity scores for high-frequency words among potential selectional equivalents. For example, you can *see* and *describe* most of the things you can *take on*, but that does not make them good selectional equivalents for either of the senses of *take on*. Effectively, these are promiscuous predicates that occur frequently with all selectors, including reliable selectors for each of the target verb's senses. Conditional probabilities for their selectors, however, are low due to their high frequencies. Normalizing the sum of minima by the sum of maxima, as in Jaccard, for example, would bring up the similarity value for high-frequency pairs such as *see* and *describe*. Without such normalization, both words in such pairs have equally low values for all nouns in their respective selector lists, which leads to a low similarity score.

Table 3. Conditional probabilities $P(s\,|\,wR)$ for the intersection of top-K selector lists for selectional equivalents of different senses of *deny*, in direct object position. A. (left) Overlapping selectors for *confirm* and *contradict*, as compared with *refuse*. B. (right) Overlapping selectors for *grant* and *refuse*, as compared with *confirm*.

	refuse	confirm	contradict		confirm	grant	refuse
report	0.000	0.018	0.006	access	0.000	0.013	0.014
claim	0.004	0.007	0.019	rights	0.001	0.015	0.002
story	0.000	0.004	0.004	permission	0.000	0.053	0.066
view	0.000	0.023	0.032	request	0.001	0.008	0.034
allegation	0.000	0.001	0.002	relief	0.000	0.012	0.009
suggestion	0.000	0.002	0.006	application	0.001	0.014	0.054
				bail	0.000	0.016	0.011

Table 4. Similarity matrix for selectional equivalents of *deny* given in Table 3. Similarity values for selectional equivalents of the same sense are underlined. Values are given for top-15 selector lists.

	refuse	grant	confirm	contradict
refuse	–	<u>0.0983</u>	0.0058	0.0064
grant	–	–	0.0059	0.0000
confirm	–	–	–	<u>0.0487</u>
contradict	–	–	–	–

There are inevitable misfires in the obtained selector lists. However, in order to compute the similarity value, we use the intersection of selector lists (cf. Eq. 1). For selectional equivalents of the same sense, this discards most of the spurious selectors chosen for each verb. Tables 3 and 4 illustrate such similarity computation for the selectional equivalents of two senses of *deny*. Table 3 (left) shows selectors chosen for *confirm* and *contradict*, the equivalents for the sense 'proclaim false'. Table 3 (right) shows selectors chosen for *grant* and *refuse*, the equivalents for the sense 'refuse to grant'. For comparison, we give conditional probabilities for the same selectors with one of the equivalents of the other sense (*refuse* and *confirm*, respectively).

The resulting similarity scores are shown in Table 4. Conditional probability values for the correctly chosen selectors cumulatively insure that the similarity between selectional equivalents of the same sense is higher than their similarity with selectional equivalents of the other sense. This similarity measure thus enables us to differentiate between senses by obtaining clusters of selectional equivalents that can then be used to identify selectors for each of the senses of the target predicate.

7. Conclusion

Sense detection is always a clustering problem, i.e. a problem of grouping together similar elements. It is a problem that is intrinsically hard to resolve distributionally, especially when argument semantics is the source of meaning differentiation. Obtaining large amounts of consistently annotated data for this task is also very difficult. In this paper, we presented an approach to this problem that relies on the notion of contextualized similarity, i.e. the fact that similarity between words should be measured with respect to a particular selection context.

The proposed method associates each of the target's senses with a cluster of selectional equivalents for that sense, with selectional equivalents represented as short contextualized vectors of reliable selectors. The resulting clusters serve to identify selectors that activate each sense, with association scores obtained for each selector indicating which sense it tends to activate. Even with certain assumptions about parallel sense distinctions and selector polysemy, we seem to be able to overcome some of the difficulties encountered by the previous attempts to address polysemy resolution. The overall results can be improved further if reliable selectors are detected in a way that does not rely to the same extent on such assumptions. The proposed approach can also be extended to multiple argument positions, since the information derived from different argument positions with respect to the likely sense of the target can be easily combined.

The output produced by the clustering algorithm can be used in a number of ways, in tasks related to sense disambiguation. The derived information about selectional properties of different senses of the target word can serve to improve the overall performance of a complete WSD or WSI system. It can also provide powerful enhancements to the lexicographic analysis tools that facilitate sense definition. For example, it can be used to create contextualized clusters of collocates in an application such as the Sketch Engine (Kilgarriff et al. 2004). In fact, examining the induced sets of selectional equivalents often reveals unexpected relationships between verbs that accept similar arguments in a given argument position. The discovered selectional equivalence relations are often impossible to predict by inspecting the data with traditional methods. This suggests that the presented technique for automated analysis of selectional properties can also be viewed as a tool for a more focused empirical study of the data. In particular, it may serve to enrich the initial models of the data, i.e. the theoretical models that are often limited to using introspective intuition and targeted corpus studies.

Address of the Author
Anna Rumshisky, Department of Computer Science, Brandeis Univeristy, Waltham, MA USA.
<arum@cs.brandeis.edu>

Notes

1 Cf. Pustejovsky (1995).
2 See Rumshisky et al. (2007) for a discussion of similar considerations for polysemous nouns.
3 These and other examples are taken, in somewhat modified form, from the British National Corpus (BNC, 2000).
4 Double square brackets are used for argument type specification, curly brackets are used for syntactic constituents, and parentheses indicate optionality. For full pattern syntax, see Pustejovsky et al. (2004).
5 For instance, out of 20 fairly frequent verbs surveyed in Rumshisky (2008), only 7 had their main sense distinctions captured in FrameNet. Only 25 out of the total 68 identified senses for these verbs had a corresponding link to a FrameNet frame. Common verbs such as *assume, claim, cut, deny, enjoy*, and *launch*, had only one out of two or three main senses represented in FrameNet.
6 Some of the similarity measures commonly used are described in Manning & Shütze (1999), Dagan (2000), Curran (2004), Lee (1999), and elsewhere.
7 Light verbs have a much higher number of senses, but we will not consider them here.
8 See, for example, Curran (2004) for a survey of different weighting schemes.
9 Association scores shown in the table are conditional probabilities *P*(selector | context).
10 pp_against is a relation between the governing verb and the head of a prepositional phrase introduced by *against*; n_modifier is a relation between a noun and a nominal modifier.
11 This graph representation is similar to the one used in literature more commonly for symmetric relations such as conjunction or apposition (Widdows & Dorow 2002) or co-occurrence within a window (Agirre et al. 2006).
12 Cluster selection, as well as steps 1 through 3 above, are described in more detail in Rumshisky et al. (2007) and Rumshisky (2008).
13 We used RASP (Briscoe & Carroll 2002) to extract grammatical relations from the British National Corpus (BNC 2000).
14 This is effectively equivalent to set-theoretic overlap used in Jaccard and Dice measures.

Bibliographical references

Agirre Eneko & Philip Edmonds (eds.) 2006. *Word Sense Disambiguation: Algorithms and Applications*. Text, Speech and Language Technology. Springer Netherlands.
Agirre Eneko & David Martínez 2004. The Basque Country University system: English and Basque tasks. In Rada Mihalcea & Philip Edmonds (eds.). Senseval-3: Third International Workshopon the Evaluation of

Systems for the Semantic Analysis of Text, July, 2004, Barcelona. Spain. 44-48.

AGIRRE Eneko, David MARTÍNEZ, Oier LOPEZ DE LACALLE & Aitor SOROA 2006. Two graph-based algorithms for state-of-the-art wsd. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. 22-23 (July, 2006). Sidney. Ausytralia. 585-593.

AGIRRE Eneko, Lluís MÀRQUEZ & Richard WICENTOWSKI (eds.) 2007. Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval2007). (June, 2007) Prague. Czech Republic, June.

ALLEGRINI Paolo, Simonetta MONTEMAGNI & Vito PIRRELLI 2003. Example-based automatic induction of semantic classes through entropic scores. *Linguistica Computazionale* 1. 1–45.

APRESJAN Jurij D. 1973. Regular polysemy. *Linguistics* 142(5). 5–32.

BNC 2000. *The British National Corpus*. The BNC Consortium. University of Oxford? http://www.natcorp.ox.ac.uk/

BRISCOE Ted & John CARROLL 2002. Robust accurate statistical annotation of general text. Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002) (May, 2002). Las Palmas, Canary Islands. 1499–1504.

CRUSE David A. 1995. Polysemy and related phenomena from a cognitive linguistic viewpoint. In DIZIER Patrick St. & Evelyne VIEGAS (eds.). *Computational Lexical Semantics*. Cambridge University Press: Cambridge. 33–49.

CURRAN James R. 2004. *From Distributional to Semantic Similarity*. University of Edinburgh PhD dissertation.

DAGAN Ido 2000. Contextual word similarity. In Dale Robert, Hermann Moisl & Harold Somers (eds.). *Handbook of Natural Language Processing*. Marcel Dekker.

DOROW Beate & Dominic WIDDOWS 2003. Discovering corpus-specific word-senses. Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics. Budapest. Hungary. 79-82.

DUNNING Ted 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19(1). 61–74.

FELLBAUM Christiane (ed.) 1998. *Wordnet: an electronic lexical database*. MIT Press.

FIRTH JOHN R. 1957. A synopsis of linguistic theory 1930-1955. In FIRTH John et al. *Studies in Linguistic Analysis*, Oxford: Blackwell. 1–32.

GALE William, Kenneth CHURCH & David YAROWSKY 1993. A method for disambiguating word senses in a large corpus. *Computers and the Humanities* 26. 415– 439.

GAMALLO Pablo, Alexandre AGUSTINI, & Gabriel P. LOPES 2005. Clustering syntactic positions with similar semantic requirements. *Computational Linguistics* 31(1). 107–145.

GREFENSTETTE Gregory 1994. *Explorations in Automatic Thesaurus Discovery*. Norwell, MA: Kluwer Academic Publishers.

HALLIDAY Michael A.K. 1973. *Explorations in the Functions of Language*. London: Edward Arnold.

HANKS Patrick & James PUSTEJOVSKY 2005. A pattern dictionary for natural language processing. *Revue Française de Linguistique Appliquée* 10(2). 63-82.

HARRIS Zelling S. (ed.) 1985. Distributional structure. In KATZ Jarrod J. (ed.). *Philosophy of Linguistics*. New York: Oxford University Press. 26-47.

HINDLE Donald 1990. Noun classification from predicate-argument structures. 28th Annual Meeting of the Association for Computational Linguistics Pittsburgh, PA. 268-275.

HOVY Edward, Mitch MARCUS, Martha PALMER, Lance RAMSHAW & Ralph WEISCHEDEL 2006. OntoNotes: The 90% solution. Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers. New York City, USA. 57-60.

KILGARRIFF Adam, Pavel RYCHLY, Pavel SMRZ, & David TUGWELL 2004. The Sketch Engine. Proceedings of Euralex. Lorient. France. 105–116.

KILGARRIFF Adam 1997. I don't believe in word senses. *Computers and the Humanities* 31. 91–113.

LEE Lillian 1999. Measures of distributional similarity. 37th Annual Meeting of the Association for Computational Linguistics. College Park, Maryland. 25–32.

LIN Dekang 1998. Automatic retrieval and clustering of similar words. Proceedings of the 17th international conference on Computational Linguistics. Montreal. Canada.

MANNING Chris & Hinrich Schütze 1999. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press.

MIHALCEA Rada, Timothy CHKLOVSKI & Adam KILGARRIFF 2004. The Senseval-3 English lexical sample task. In MIHALCEA Rada & Phil EDMONDS (eds.). Senseval-3: Third International Workshopon the Evaluation of Systems for the Semantic Analysis of Text. (July, 2004). Barcelona. Spain, July. 25-28.

MILLER George & Walter G. CHARLES 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6(1). 1-28.

NAVIGLI Roberto, Kenneth LITKOWSKI & Orin HARGRAVES 2007. Semeval-2007 task 07: Coarse-grained English all-words task. Proceedings ofthe Fourth International Workshop on Semantic Evaluations (SemEval-2007). (June, 2007). Prague. Czech Republic. 30-35.

NAVIGLI Roberto 2006. Meaningful clustering of senses helps boost word sense disambiguation performance. Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics. Sydney, Australia. 105–112.

PALMER Martha, Hoa Trang DANG & Christiane FELLBAUM to appear. Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Journal of Natural Language Engineering*.

PANTEL Patrick & Dekang LIN 2002. Discovering word senses from text. Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 613-619.

PRADHAN Sameer S., Edward LOPER, Dmitry DLIGACH & Martha PALMER 2007. Semeval-2007 task-17: English lexical sample, srl and all words.

Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007). (June, 2007). Prague. Czech Republic. 87–92.

PREISS Judita & David YAROWSKY (eds.) 2001. Proceedings of the Second International Workshop on Evaluating WSD Systems (Senseval 2). ACL2002/EACL2001.

PUSTEJOVSKY James, Patrick HANKS, & Anna RUMSHISKY 2004. Automated Induction of Sense in Context. Proceedings of the Fifth International Workshop on Linguistically Interpreted Corpora (August 29, 2004). (COLING 2004). Geneva. Switzerland. 924–931.

PUSTEJOVSKY James 1995. *Generative Lexicon*. Cambridge, MA: MIT Press.

RESNIK Philip 1996. Selectional constraints: An information-theoretic model and its computational realization. Cognition 61. 127–159.

RUMSHISKY Anna, Victor A. GRINBERG & James PUSTEJOVSKY 2007. Detecting Selectional Behavior of Complex Types in Text. In BOUILLON Pierrette, Laurence DANLOS & Kyoko KANZAKI (eds.). Fourth International Workshop on Generative Approaches to the Lexicon. Paris. France.

RUMSHISKY Anna 2008. *Verbal Polysemy Resolution through Contextualized Clustering of Arguments*. Brandeis University Ph.D.

RUPPENHOFER Josef, Michael ELLSWORTH, Miriam PETRUCK, Christopher R. JOHNSON & Jan SCHEFFCZYK 2006. *FrameNet II: Extended Theory and Practice*. Berkeley: FrameNet Project.

H. SCHÜTZE 1998. Automatic word sense discrimination. *Computational Linguistics* 24(1). 97–123.

SINCLAIR John M. & Patrick HANKS 1987. Glasgow: Harper Collins (First edition 1987, Collins Cobuild English Language Dictionary). *Collins Cobuild Advanced Learner's English Dictionary*.

SNYDER Benjamin & Martha PALMER 2004. The english all-words task. In Rada MIHALCEA & Philip EDMONDS (eds.). Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text. (July, 2004). Barcelona. 41–43.

VELLDAL Erik 2005. A fuzzy clustering approach to word sense discrimination. Proceedings of the 7th International conference on Terminology and Knowledge Engineering. Copenhagen. Denmark. 279-292.

WIDDOWS Dominic & Beate DOROW 2002. A graph model for unsupervised lexical acquisition. 19th International Conference on Computational Linguistics. (August, 2002). Taipei. Taiwan. 1093-1099.