Inferring Semantic Similarity from Distributional Evidence: an Analogy-based Approach to Word Sense Disambiguation*

Stefano Federici[1]
Simonetta Montemagni[1]
Vito Pirrelli[2]

[1]**Par.O.La sas, Pisa, ITALY**
[2]**Istituto di Linguistica Computazionale CNR, Pisa, ITALY**

Abstract

The paper describes an analogy-based measure of word-sense proximity grounded on distributional evidence in typical contexts, and illustrates a computational system which makes use of this measure for purposes of lexical disambiguation. Experimental results show that word-sense analogy based on contexts of use compares favourably with classical word-sense similarity defined in terms of thesaural proximity.

1 Introduction

Sense disambiguation of a given word occurrence in a specific context (hereafter WSD) requires appeal to a wide typology of cues, ranging from syntactic subcategorization to lexico-semantic information and subject domain. In this paper we will focus on the use of lexico-semantic information, and will try to tackle the related problem of measuring the semantic similarity between the surrounding context of the word to be disambiguated and typical patterns of use of that word in a dictionary database. In the literature, semantic similarity is usually assessed with reference to a hierarchically structured thesaurus (e.g. WordNet, [Miller, 1990]). The goal of the paper is to investigate an alternative way of measuring semantic similarity, based on distributional evidence, and to show that this evidence can reliably be used to disambiguate words in context. To this end, we will make use of textual and lexical resources of Italian: nonetheless we are convinced that the general point made in this paper has a cross-linguistic validity.

2 Semantic Similarity and WSD

Most methods proposed in the literature for establishing the semantic similarity of words try to map a given word in context onto the set of known usages of that word in a dictionary database: thesaural information is used as a yardstick for measuring the semantic proximity between known patterns of use and the context to be disambiguated. Eventually, the sense supported by those patterns which are semantically closer to the context in question is selected as the most likely one (see, among others, [Dolan, 1994], [Resnik, 1995a, 1995b], [Agirre and Rigau, 1996], [Sanfilippo, 1997]).

Suppose that one wants to disambiguate the sense of *accendere* in the verb-object pair *accendere-televisione* 'switch_on-tv'. The relevant sense of *accendere* can be inferred on the basis of known examples such as *accendere_2-radio* 'switch_on-radio': this inference is supported by any semantic hierarchy where both radio and television are specified for the same hyperonym, e.g. 'device', whether immediate or not.

Thesaural relationships such as hyperonymy and synonymy, however, do not always capture the dimension of similarity relevant to the context in question. Consider the verb *accendere* in the context *accendere-pipa* 'light-pipe'. The table below contains typical objects of two senses of *accendere*, 'light' (sense 1) and 'switch_on' (sense 2) as they are attested in the Collins Italian-English Dictionary [1985], together with the objects' corresponding hyperonyms according to a monolingual Italian dictionary [Garzanti, 1984].

* The work reported in this paper was jointly carried out by the authors in the framework of the SPARKLE (Shallow PARsing and Knowledge extraction for Language Engineering) project (LE-2111). For the specific concerns of the Italian Academy only, S Federici is responsible for sections 3.2, 3.4 and 3.5, S. Montemagni for 2, 3.3 and 4, and V. Pirrelli for 1, 3.1 and 5.

Table

verb sense	object	1st hyper.	nth hyper.	nth+1 hyper.
accendere_1	*sigaretta* 'cigarette'	small roll	. > artifact	object
accendere_1	*candela* 'candle'	lamp	. > artifact	object
accendere_1	*fiammifero* 'match'	small stick	> artifact	object
accendere_1	*camino* 'fireplace'	hollow	. > artifact	object
accendere_2	*motore* 'engine'	device	> artifact	object
accendere_2	*lampada* 'lamp'	source of illumination	.. > artifact	object
accendere_2	*radio* 'radio'	receiver	> artifact	object

The word *pipa*, which Garzanti describes as a smoking tool, does not match any of the immediate hyperonyms of the typical objects of *accendere_1* and *accendere_2* in **Fout! Onbekende schakeloptie-instructie.**. By looking further up in the semantic hierarchy, some similarities are indeed found, but they are based on too general semantic features to be of avail for discriminating among senses 1 and 2 of *accendere*.

We suggest that, for *accendere-pipa* to be understood in the appropriate sense, namely *accendere_1* as in *accendere_1-sigaretta* 'light-cigarette', semantic proximity need be computed on different grounds. The relevant similarity wich links pipes and cigarettes in this specific context relates to their both being typically smoked objects, a fact which is orthogonal to their general semantic class and can be captured on a distributional basis: *pipa* and *sigaretta* are distributionally equivalent relative to the same verb sense, i.e. they both occur as typical objects of the verb *fumare* 'smoke'. Distributional equivalence correlates with semantic similarity under the assumption that nouns which bear the same syntactic relation to the same verb sense are part of a semantically coherent class. It turns out that, in examples such as *accendere-pipa*, distributionally-based semantic similarities can permit more appropriate sense assignments which are specifically tailored to the context to be disambiguated. Observe further that also similarities commonly captured on the basis of thesaural information, as in the case of *radio* and *televisione* above, can in principle be inferred from distributional evidence through relevant contexts of use (e.g. *spegnere-radio* 'switch_off-radio' and *spegnere-televisione* 'switch_off-tv' in the example at hand).

Summing up, we contend that thesaural relationships capture only some of the various dimensions of word sense analogy which appear to play a relevant role in the disambiugation of word co-occurrence patterns. In fact, while thesaural relationships are defined out of context once and for all, effective analogies are to be tailored to the specific contextual pattern to be disambiguated. We showed how this can be attained on the basis of distributional evidence.

3 SENSE: a distributionally-based WSD system

SENSE (Self-Expanding linguistic kNowledge-base for Sense Elicitation) is a specialised version of a general purpose language-learning system ([Federici and Pirrelli, 1994]; [Federici *et al.*, 1996a]; [Montemagni *et al.*, 1996]) for carrying out WSD on the basis of distributional evidence.

SENSE's inferential routine requires:

i) a structured data set of known word co-occurrence patterns (WCPs) constituting an Example Base (EB);
ii) a target context to be disambiguated (TC);
iii) a best-analogue(s) function (BAF) projecting TC onto EB for the best analogue(s) to be selected and thus the most likely senses to be identified.

3.1 Internal architecture of EB

Word co-occurrence patterns

WCPs are modelled here as consisting of an input and an output level of representation. At the input level, each element of the pattern is described by a set of features which are expected to be of some use for WSD: lemma, part of speech and morpho-syntactic properties (such as the syntactic function of nouns with respect to the verb). The output representation simply consists in the expected answer, i.e. the sense of each element of the pattern in the described context. An example of this type of linguistic object, illustrating the pattern *fumare_1-sigaretta_1/O* 'smoke-sigarette', is given in **Fout! Onbekende schakeloptie-instructie.**:

Table

	fumare_1-sigaretta_1/O	
input	fumare	sigaretta
	verb	noun
		object
output	fumare_1	sigaretta_1

The input representation is a list of sets of atomic units; each feature set (which is assigned a single column in the table) describes a distinct element of the pattern. In output, the list of atomic units "fumare_1" and "sigaretta_1" indicates the senses of the elements in the specific context. Elements in the input and output lists are conventionally ordered.

In the current version of Italian EB used for our purposes, WCPs are verb-noun pairs where the relation of the noun to the verb is either subject or object. This presupposes a preliminary stage of morpho-syntactic parsing [Montemagni, 1995]: co-occurrence patterns abstract away from actual word forms and are augmented with information about grammatical relations. Note that although availability of pre-processed input makes word sense disambiguation simpler and more accurate, it is in no way a necessary precondition for the task to be carried out.

Pairwise analogies

The Italian EB consists of WCPs of the type illustrated in **Fout! Onbekende schakeloptie-instructie.** above. Note however that they are not used as such; rather they form part of a distributed network[1] which is constructed so as to i) factor out the optimal set of analogies shared by all WCPs in EB, and ii) link the found analogies with their corresponding complements relative to the full WCPs (so-called differing parts). To make this picture more concrete, let us consider some simple examples.

Given a pair of word co-occurrence patterns wcp_1 and wcp_2, they are judged to be analogous if they share some representation units at both input and output levels. Any shared collection of units of both levels is referred to as an analogical core (or simply core, written $wcp_1 \cap wcp_2$). Suppose that wcp_1 and wcp_2 are *fumare*_1-*sigaretta*_1/O and *fumare*_1-*pipa*_1/O 'smoke-pipe' respectively, defined as in **Fout! Onbekende schakeloptie-instructie.** above and **Fout! Onbekende schakeloptie-instructie.** below.

Table

	*fumare*_1-*pipa*_1/O	
input	fumare	pipa
	verb	noun
		object
output	fumare_1	pipa_1

Their core is identified by a function (MF) mapping one set of units in **Fout! Onbekende schakeloptie-instructie.** onto one set of units in **Fout! Onbekende schakeloptie-instructie.** through the identity relation. MF is order-sensitive, so that only sets which take the same relative order in the lists are mapped onto each other. **Fout! Onbekende schakeloptie-instructie.** gives a possible result of this operation in the leftmost box headed by $wcp_1 \cap wcp_2$. The core in question is a verb-noun pair where the noun element is specified only at the input level, for a subset of the features describing the noun elements in the compared patterns, while nothing being said as to the possible sense interpretation of the noun. Nonetheless, the information about the noun conveyed by the core, namely its syntactic relation to the verb, is part of the knowledge supporting the interpretation of the verb as *fumare*_1: i.e. the verb in this reading is used transitively.

Table

	$wcp_1 \cap wcp_2$		$wcp_1 - wcp_2$	$wcp_2 - wcp_1$
input	fumare		sigaretta	pipa
	verb	noun		
		object		
output	fumare_1		sigaretta_1	pipa_1

The complements of the core relative to wcp_1 and wcp_2 designate those units which are specific to the compared objects: they constitute the so-called differing parts, illustrated in **Fout! Onbekende schakeloptie-instructie.** in the columns headed by $wcp_1 - wcp_2$ and $wcp_2 - wcp_1$ respectively. They contain information about the lexical fillers of the noun slots of the patterns.

Network structure of EB

Cores and remaining parts are always anchored to a given pair of linguistic objects: in fact, cores cannot be extracted either from existing cores or from existing differing parts. When more than one pair of WPCs is considered, it may turn out that what is a core relative to a given pair is a remaining part relative to another pair. Suppose that MF maps *fumare*_1-*sigaretta*_1/O (wcp_1) onto *accendere*_1-*sigaretta*_1/O (wcp_3). One of the possible results of this mapping is shown in **Fout! Onbekende schakeloptie-instructie.** below:

Table

	$wcp_3 - wcp_1$		$wcp_1 - wcp_3$		$wcp_1 \cap wcp_3$
input	accendere		fumare		sigaretta
	verb	noun	verb	noun	
		object		object	
output	accendere_1		fumare_1		sigaretta_1

Comparison of cores and remaining parts in **Fout! Onbekende schakeloptie-instructie.** and **Fout! Onbekende schakeloptie-instructie.** above shows that one of the remaining parts relative to wcp_1 and wcp_2 (namely $wcp_1 - wcp_2$) is identical to the core relative to wcp_1 and wcp_3 ($wcp_1 \cap wcp_3$).

The informational content of **Fout! Onbekende schakeloptie-instructie.** and **Fout! Onbekende schakeloptie-instructie.** can be represented conveniently through the graph in **Fout! Onbekende schakeloptie-instructie.**.

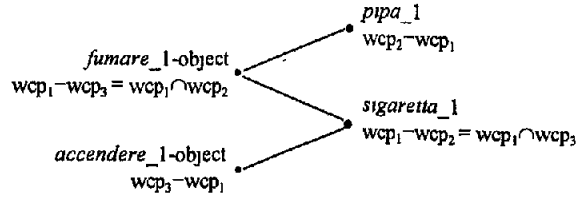

Figure An analogical family

The graph represents cores and remaining parts as connected nodes, each accompanied by a mnemonic label. For example *sigaretta*_1 corresponds to $wcp_1 - wcp_2 = wcp_1 \cap wcp_3$. An (unoriented) arc connecting two nodes expresses their "complementarity", i.e. the intuitive notion that the two connected nodes, taken together, cover an attested WCP in its entirety. For instance, *fumare*_1-object is connected with *sigaretta*_1 since they form together an attested WCP, namely *fumare*_1-*sigaretta*_1/O. By contrast, no direct connection is ob-

[1] In the current version of SENSE a (partial) network structure is built from scratch every time a new TC is presented to the system However, for the sake of clarity, in what follows we illustrate the working of our system as though the network structure were built during the acquisition of EB. See [Federici *et al*, 1996b, p.393] for a discussion of the two alternatives.

served between *accendere*_1-object and *pipa*_1, to signify that no corresponding pattern is attested in EB. Remaining parts which are connected with the same core are said to be contrastive, since, by replacing one with the other, different WCPs are obtained. A graph like the one in **Fout! Onbekende schakeloptie-instructie.** represents in our terms an analogical family (AF). Clearly, far more extended AFs than the one in **Fout! Onbekende schakeloptie-instructie.** can be found.

Among the WPCs of the AF in **Fout! Onbekende schakeloptie-instructie.**, *fumare*_1-*sigaretta*_1/O is the only one which is made up out of two cores, namely $wcp_1 \cap wcp_3$ and $wcp_1 \cap wcp_3$. Due to its pivotal position in the graph, it is some times useful to refer to it as the "hook pattern", or more simply "hook", of the AF in question. Accordingly, we will call the noun collocate of a hook, i.e. *sigaretta*_1 in the example at hand, "hook noun", and the corresponding verb, i.e. *fumare*_1, "hook verb". Note further that the hook noun *sigaretta*_1 is functional to establishing a kinship between the verb senses *fumare*_1 and *accendere*_1, since it denotes a non-empty intersection between typical patterns of their use.

3.2 The Best-analogue(s) Function

Unlike linguistic objects in EB, which are specified for two representation levels (input and output), a Target Context (TC) is specified at the input level only, since its sense is precisely what the system has to predict on the basis of the available knowledge.

This prediction is carried out through operation of the best-analogue(s) function (BAF) which projects TC onto EB, searching for TC's best candidate analogue(s). BAF uses the notion of distributionally-driven word-sense analogy developed in the previous pages, and can be informally described through the following steps:

a) if EB contains a pattern wcp_i which fully matches TC at the input level, then wcp_i is the best analogue and its output is ranked first in the list of available answers; note that this step does not stop SENSE from continuing its search;
b) if EB contains a single AF such that two of AF's nodes together cover TC's input representation in its entirety, the output representations associated with the matching nodes is added to the list of available answers with a ranking score, gauged as a function of type and quantity of supporting evidence (see below for more detail);
c) if steps a) and b) yield no result, no output is provided by SENSE.

BAF at work

Let us look at some interesting cases of BAF at work. Note that all examples considered in this paper are representative of real test suites of SENSE, and the assumed knowledge in EB reflects the current status of an actual data base acquired from typical examples of use within verb entries of the Collins Italian-English dictionary [1985].

Suppose that SENSE has to assign a verb sense in the target context *accendere*_?-*pipa*_1/O 'light-pipe'. The context being not attested in EB, TC is projected onto EB's network, for a relevant AF to be found. The AF in **Fout! Onbekende schakeloptie-instructie.** above is a good instance of such a relevant family, since it contains two nodes, namely *accendere*_1-object and *pipa*_1, which fully cover TC's input. Step a) having failed, the two nodes in question are not directly connected; nonetheless, their belonging to the same family means that there exists a continuous path of complementarity arcs joining the two. This continuity allows SENSE to hypothesize an arc directly connecting *accendere*_1-object with *pipa*_1 (represented as a dashed line in **Fout! Onbekende schakeloptie-instructie.** below):

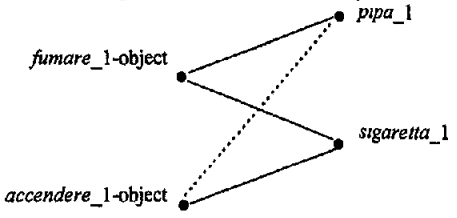

Figure A reconstructed connection

i.e. the co-occurrence pattern *accendere*_1-*pipa*_1/O can be reconstructed on the basis of the available distributional evidence, and supports the interpretation *accendere*_1.

To sum up, SENSE identifies a distributional similarity between *accendere*_1 and *fumare*_1: this similarity is based on the fact that cigarettes can both be lit and smoked. This triggers the analogy-based inference that pipes, besides being smoked, can also be lit, thus supporting the interpretation of *accendere*_1

3.3 Constraints on distributionally-based WSD

In the example illustrated above, nouns stand in the same syntactic relation to the verbs. However, it is often the case that clusters of nouns which function as the object of a given verb can function as typical subjects of other, somehow related, verbs. If this sort of systematic subject/object alternation is taken into account, the inferential power of distributionally-based WSD may increase considerably, as shown by the following examples.

Consider the TC *attaccare*_?-*fotografia*_1/O 'hang_up-photograph'. EB contains three different senses of *attaccare*, each attested with the following sets of noun collocates:

- *attaccare*_1-[*francobollo*_1/O, *manifesto*_1/O, *quadro*_1/O]
  'hang_up'-['stamp/O', 'poster/O', 'painting/O']
- *attaccare*_2-[*discorso*_1/O]
  'start'-['conversation/O']
- *attaccare*_4-[*moda*_1/S]
  'catch_on'-['fashion/S']

No one of the noun collocates listed above happens to be attested in EB as an object of verbs which also combine

with *fotografia* as an object. However, if the restriction that relevant nouns must stand in the same relation to the predicate is relaxed, then relevant distributional evidence can in fact be found in EB. *Fotografia* and *quadro* 'painting' are both attested as typical subjects of the verb *rappresentare_1*, a fact which can be interpreted in terms of Pustejovsky's telic role [Pustejovsky 1995], since both nouns are normally used to "show something". Furthermore, *quadro* is also attested as a typical object of the verb *attaccare_1*; on this basis, it can reasonably be supposed that also *fotografia*, when co-occurring as an object of *attaccare*, points to the sense *attaccare_1*.

Inferences based on AFs involving asymmetric syntactic dependencies permit to exploit the data contained in EB to the full. Moreover, the procedure becomes essential for generalising over cases of so-called valency alternation. Consider the causative-inchoative alternation, which involves two argument structures, a transitive and an intransitive one: a verb such as *aumentare* 'increase' can be used in a sentence like *la Fiat ha aumentato gli stipendi agli operai* 'Fiat increased salaries to workers', where *stipendio* is the object of the verb, and in a sentence like *gli stipendi aumentarono inaspettatamente* 'salaries increased unexpectedly', where *stipendio* is the subject. In the literature, the theoretical issue of whether alternating argument structures of the same verb should be associated with a unique sense or with different senses of that verb is still open. In practice, lexicographers' approaches vary considerably, depending on factors such as the dictionary's internal structure or main practical purpose: for instance, in bilingual dictionaries different but alternating argument structures often give rise to different senses, due to differences in their translation. Whatever approach is adopted by the lexicographer, however, SENSE is capable of identifying a sense alternation induced by an alternation of argument structure, or, alternatively, of recognising two different argument structures as related to the same verb sense, thanks to its ability to deal with asymmetric syntactic dependencies in EB.

To sum up, word sense disambiguation with verb-noun pairs involving asymmetric dependencies is more effective than when only contexts with symmetric dependencies are considered. This procedure is particularly crucial for verbs alternating between different arguments structures.

3.4 Beyond attested evidence

SENSE's inferential routine can go beyond attested evidence; in fact, the presence of an attested pattern which matches exactly TC's input does not prevent the system from exploring other hypotheses. This flexibility is often useful: when sense distinctions are fine grained and data in EB are sparse, distributional criteria get too coarse grained to be able to point to a unique sense interpretation.

Consider, for example, *battere_?-mano_1*/O 'hit-hand': in EB, this pattern is attested with the sense of clapping, as an instance of beating body parts with a regular rhythm (*battere_3*). However, there is at least another sense of *battere* which is appropriate in the context considered, namely *battere_1*, understood under the more general sense of hitting someone or something. In cases like this one, SENSE "ambiguates" the verb-noun pair received in input, by finding out other plausible sense assignments besides the one attested in EB. As a consequence, SENSE outputs more than one sense interpretation, while ranking the attested interpretation first. Identification of alternative sense assignments, although with lower ranking, comes in handy when the expected TC interpretation is not the attested one. This is reasonable, we believe, since WSD is often a matter of suggesting a set of more or less plausible interpretations in context rather than asserting one interpretation only; by taking attested evidence (no matter how representative) at face value one would wrongly ignore the common fact that, even in real usages, a target context can in fact be understood in more than one way.

3.5 Ranking multiple disambiguation results

As just shown, distributionally-based word sense disambiguation does not always make the system converge on a unique interpretation. This situation typically occurs when different senses of a word are close in meaning, and this closeness is reflected by their co-occurrence with distributionally similar if not identical words. When more than one sense interpretation appears to be plausible, different strategies can be followed in order to rank them from more to less likely. When the set of plausible interpretations includes a directly attested one, then the latter is always ranked first. Ranking of inferred interpretations needs to take into account a number of different factors.

As a first approximation, different sense interpretations can be ranked according to the number of AFs supporting them. Suppose that SENSE has to assign a verb sense in *accarezzare_?-speranza_1*/O 'toy_with-hope'. Both possible sense interpretations of *accarezzare* (i.e. *accarezzare_1* 'stroke' and *accarezzare_2* 'toy_with') are supported. In EB, the interpretation *accarezzare_1* is supported by one AF only, which includes the pattern *perdere_1-capello_1*/O 'lose-hair'. On the other hand *accarezzare_2* is supported by four AFs, each containing the following hooks:

1. *abbandonare_4-progetto_1*/O 'give_up-project'
2. *cullare_1-idea_1*/O 'cherish-idea'
3. *nascere_2-idea_1*/S 'be_born-idea'
4. *naufragare_1-progetto_1*/S 'fall_through-project'

Hence, *accarezzare_2* gets score 4 and wins out over *accarezzare_1* which scores 1.

The sheer number of supporting AFs, however, is too gross a criterion when used on its own. Consider the target *affluire_?-acqua_1*/S 'flow-water'. Here, the contextually more appropriate sense *affluire_1* 'flow' is supported by three AFs, while *affluire_2* 'pour_in' is pointed to by five different AFs:

affluire_1

1. *intorbidare_1-liquido_1*/O 'cloud-liquid'
2. *penetrare_2-liquido_1*/S 'percolate-liquid'
3. *versare_2-liquido_1*/O 'pour-liquid'

affluire_2
1. *confluire_1-persona_1*/S 'join-person'
2. *gettare_1-persona_1*/O 'rush_in-person'
3. *imbarcare_1-merce_1*/O 'ship-goods'
4. *insinuarsi_3-persona_1*/S 'creep_into-person'
5. *ristagnare_1-persona_1*/S 'lag_person'

Nonetheless, SENSE could be "persuaded" to prefer the correct interpretation if also the typology of supporting evidence is taken into account. Intuitively, preference has to be given to more specific supporting semantic evidence over semantically vaguer one. In our terms, this means that supporting AFs which contain a more specific hook noun should carry more weight for WSD than AFs containing vaguer hook nouns. Usually, generality of a word is measured by referring to a semantic hierarchy. In this context we have used frequency of word occurrence in EB as a convenient measure of "generality/specificity" of a word: the more often a hook noun occurs as a subject/object of differents verbs in EB, the more general it can be considered. Note that EB contains only WCP types, so that word counting here is significantly different from counting token frequencies in a real text; type frequency appears to point more decisively to the general structure of lexical competence, rather than to distributional effects in language performance. On this basis, each relevant AF is assigned a specificity score, equal to the inverse ratio of the number of times its hook noun occurs in EB. The ranking score of a given sense interpretation S is then the sum of the specificity scores of all AFs = { AF_1, AF_2, ..., AF_n } supporting it:

$$S_{spec\_score} = Spec(AF_1) + Spec(AF_2) + \ldots + Spec(AF_n)$$

where $Spec(AF_i)$= 1/type-frequency(hook_noun).

In the light of this score, ranking of the senses of *affluire* is reversed: the best disambiguation hypothesis is now *affluire_1* (ranking score 0.281046), against *affluire_2* whose ranking score 0.069598 is significantly lower. The hook noun supporting the sense *affluire_1* is *liquido* 'liquid', whose specificity score is 0.111111 when used as an object and 0.058824 when used as a subject. By contrast, the same score is significantly lower in the cases supporting the other sense: 0.007194 for *persona_1*/O and 0.005650 for *persona_1*/S; 0.045455 for *merce_1*/O.

The specificity score tends to overrate very specific analogies, that is analogies supported by analogical families with highly idiosynractic collocates, over more general analogies. To counterbalance this bias, another ranking factor, called the "coverage" score, can usefully be exploited in our context. For each available sense interpretation of TC attested in EB, we count how many of its collocates occur as hook nouns of all AFs supporting that sense. Note that, for an AF to support a certain verb sense, it has to contain as a hook noun a collocate of the verb sense S in question. We then assign to S a coverage score $S_{coverage\_score}$, which is proportional to the number of shared collocates:

$$S_{coverage\_score} = \#\_noun\_collocate(AF(S))/\ \#\_noun\_collocate(S)$$

where '#_noun_collocate(AF(S))' reads "cardinality of the noun collocates of the AFs supporting S", and '#_noun_collocate(S)' reads "cardinality of the noun collocates of S". The bigger this score, the more widely supported the corresponding sense interpretation in EB. This follows quite naturally in an analogy-based perspective, since, intuitively, two verb senses are considered more similar if they have more collocates in common. Eventually, this score is combined with the other scores considered above to yield a final ranking score:

$$S1_{Overall\_ranking\_score} = S_{spec\_score} \times S_{coverage\_score}$$

To give a concrete example, assume that SENSE has to interpret the pattern *accostare_?-qualcuno_1*/O 'approach-somebody'. *Accostare* is attested in EB in three different senses: *accostare_1* 'bring_near' with words like chair, object and ladder, among its typical objects; *accostare_2* 'approach' with person as typical object; *accostare_3* 'set_ajar' said of shutter and door. If the coverage score is not considered, the ranking would be *accostare_3* (0.428571), *accostare_1* (0.316417), *accostare_2* (0.122302) the latter being the appropriate sense in this context. Intuitively this is due to the fact, that, for example, the AFs supporting *accostare_3* all exhibit one hook noun only, namely *porta*, which nonetheless contributes a high specificity score, due to its poor type-frequency in EB. Yet, if the coverage score is taken into account, the ranking becomes *accostare_2* (2.079136), *accostare_1* (1.582087), *accostare_3* (0.428571), with the appropriate sense ranked first.

4 SENSE: experimental results

Experiments have been carried out with an EB of 8,153 distinct verb-noun patterns (2,488 verb-subject, 5,665 verb-object) automatically extracted from the whole set of verb entries of the Collins bilingual Italian-English dictionary [Montemagni, 1995]. In these patterns only verbs are disambiguated as to their sense, whereas nouns are assigned all possible senses. These patterns exemplify 3,359 different verb senses, each illustrated, on average, through 2.42 patterns. In **Fout! Onbekende schakeloptie-instructie.** below, verb senses are ranked per number of exemplifying patterns:

Table

n of patterns	verb senses	
10-15	21	0 6%
9-6	188	5 6%
2-5	1874	55 8%
1	1276	38%
total	3359	100%

Senses which are attested in ten or more patterns are a negligible part of EB; actually, most verbs are illustrated by means of a number of patterns ranging between 2 and 5. Finally, a considerable group of verb senses is attested only once. Note that this does not stop SENSE from recognising them in unseen contexts; e.g. in EB there is

only one pattern exemplifying the verb sense *abbassare_3* 'reduce' (namely, *abbassare_3-prezzo*/O), but this does not prevent SENSE from recognising it in target contexts such as *abbassare_?-stipendio*/O.

SENSE's performance has been tested on a corpus of 150 TCs randomly extracted from unrestricted texts. Patterns which already occur in EB were excluded from the test corpus since we wanted to focus on the reliability of inferences based on distributional evidence, rather than on EB's statistical representativity. The results of this experiment are reported below:

Table

	Overall	Polysemous
RECALL	79 3%	66 3%
PRECISION	89 9%	80 4%

Figures in the first column refer to both polysemous and monosemic verbs; here, recall and precision are high and refer to the topmost sense in the ranking only. In the second column, recall and precision are relative to polysemous verbs only, and in spite of an obvious decrease compare well with related work carried out with different methods (see, for instance, [Agirre and Rigau, 1996]), and are in fact very promising if one considers the comparatively small size of EB, and that only part of its attested words are semantically disambiguated.

5 Concluding remarks

In this paper we described a WSD system which uses a notion of semantic similarity based on distributional evidence. Preliminary results look promising.

The described measure of semantic similarity offers significant advantages compared with methods where word similarity is evaluated either in statistical terms, ultimately based on token frequency, or through reference to a hierarchically structured thesaurus. First, good results are achieved with small quantities of data, part of which are not even semantically disambiguated. Second, the suggested measure is sensitive to similarities which are relevant to the context being disambiguated, thus overcoming one of the major drawbacks of fixed decontextualised semantic hierarchies.

On a more practical front, this measure was evaluated as an integral part of the disambiguation strategy of SENSE, whose main advantages over other WSD systems can be summarised as follows:

- SENSE does not take attested evidence at face value but always entertains other hypotheses;
- SENSE's inferences are not restricted to contexts which exhibit symmetric syntactic dependencies, but also exploit alternations in argument surface realisation with semantically related verbs;
- SENSE is sensitive to the semantic generality/specificity of supporting evidence.

References

[Agirre and Rigau, 1996] E. Agirre, G. Rigau. Word Sense Disambiguation using Conceptual Density. *Proceedings of COLING-96*, Copenhagen, Denmark, pp. 16--22, 1996.

[Collins, 1985] Collins Giunti Marzocco. *English-Italian Italian-English Dictionary*. Collins Giunti Marzocco, London Firenze, 1985.

[Dolan, 1994] W. Dolan. Word Sense Ambiguation: Clustering related Senses. *Proceedings of COLING-94*, Kyoto, Japan, pp. 712--716, 1994.

[Federici and Pirrelli, 1994] S. Federici, V. Pirrelli. Linguistic Analogy as a Computable Process. *Proceedings of NeMLaP*, Manchester, UK, pp. 8--14, 1994.

[Federici *et al.*, 1996a] S. Federici, S. Montemagni, V. Pirrelli. Analogy and Relevance: Paradigmatic Networks as Filtering Devices. *Proceedings of NeMLaP*, Ankara, Turkey, pp. 13--24, 1996.

[Federici *et al.*, 1996b] S. Federici, V. Pirrelli, F. Yvon. A dynamic Approach to Pardigm-driven Analogy. in S. Wermter, E. Riloff, G. Scheler (Eds.) *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, pp.385-398, Springer, 1996.

[Garzanti, 1984] Garzanti. *Il Nuovo Dizionario Italiano Garzanti*. Garzanti, Milano, 1984.

[Miller, 1990] G. Miller. *Five Papers on WordNet*. Special Issue of *International Journal of Lexicography*, 3(4), 1990.

[Montemagni, 1995] S. Montemagni. *Subject and Object in Italian Sentence Processing*, PhD Dissertation, UMIST, Manchester, UK, 1995.

[Montemagni *et al.*, 1996] S. Montemagni, S. Federici, V. Pirrelli. Resolving syntactic ambiguities with lexico-semantic patterns: an analogy-based approach *Proceedings of COLING-96*, Copenhagen, August 1996, pp. 376--381, 1996.

[Pustejovsky, 1995] J. Pustejovsky. *The Generative Lexicon*. The MIT Press, Cambridge, Massachusetts, 1995.

[Resnik, 1995a] P. Resnik. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. *Proceedings of IJCAI-95*, 1995.

[Resnik, 1995b] P. Resnik. Disambiguating noun groupings with respect to WordNet senses. *Proceedings of 3rd Workshop on very large corpora*, Association for Computational Linguistics, 1995.

[Sanfilippo, 1997] A. Sanfilippo. *Using Semantic Similarity to Acquire Cooccurrence Restrictions from Cor-*

pora, SPARKLE Project (LE 2111), Working paper n.12, 1997.