Augmenting WordNet-like lexical resources with distributional evidence. An application-oriented perspective*

Simonetta Montemagni, Vito Pirrelli
Istituto di Linguistica Computazionale, CNR
Via della Faggiola 32, Pisa, ITALY
e-mail: {simo,vito}@ilc.pi.cnr.it

Abstract

The paper deals with the issue of how and to what extent WordNet-like resources provide the necessary information for an assessment of semantic similarity which is useful for practical applications. The general point is made that taxonomical information should be complemented with distributional evidence. The claim is substantiated through experimental data and an illustration of a word sense disambiguation system (SENSE) capable of using contextually-relevant semantic similarity.

1. Introduction

Assessment of semantic similarity has proved to be be essential for a variety of Natural Language Processing (NLP) tasks, including syntactic disambiguation (either structural or functional), word sense disambiguation, selection of appropriate translation equivalent, assessment of lexical cohesion in texts for automatic summarisation, query expansion and document indexing in Information Retrieval.

Typically, the semantic similarity between words is computed on the basis of taxonomical relationships such as hyperonymy. Given two word senses W_1 and W_2, their similarity is captured as a function of their belonging to more general semantic classes. The approach presupposes prior availability of independent hierarchically-structured repositories of lexico-semantic information such as WordNet. An interesting issue here is to evaluate how useful this type of resource is in capturing semantic similarity at the desirable level of granularity, given the requirements of the abovelisted applications.

As a general comment, the taxonomical approach to semantic similarity tends to neglect the role of linguistic context as a perspectivising factor affecting the perception of a semantic similarity between any two words considered. There is substantial experimental evidence supporting the view that human similarity judgements are affected by the pressure of contextual factors (see, among others, Goldstone et al. 1997): intuitively, while *candle* and *barbecue* would score poorly on semantic proximity if considered independently of their use in context, their occurence in expressions such as *light a candle*, *light a barbecue* would immediately throw in relief a (possibly weak, but nonetheless contextually relevant) semantic association between the two, established by their connection with the process of burning. This association is relevant insofar as it plays a role in carving out the set of plausible objects of the verb *light*. Taxonomies are not in principle incapable of capturing cross-classifications like those based on relational or role properties such as "being a product" or "being a typical object of event/process". There are allowances in the latest Wordnet version (1.6) for defining pointers from each concept to - say - nouns representing its parts, or from nouns to verbs to represent functions etc., although the latter are not actually implemented yet. Nonetheless, it is not obvious how many of these cross-classificatory dimensions should be overlaid on a taxonomy to attain the desirable level of context-sensitivity required by real applications. From an application-oriented perspective, there is the further problem of how it is possible to regiment their role and relevance as a function of context variation.

As a somewhat radical alternative to taxonomical relationships, other ways of measuring semantic similarity based on distributional evidence have been put forward in the literature (see, among others, Brown et al. 1991, Gale et al. 1992, Pereira and Tishby 1992), which emphasise the role played by context in this game. These approaches compute the semantic similarity between W_1 and W_2 on the basis of the extent to which W_1/W_2's average contexts of use overlap. Here, the context is generally defined as an *n*-word window centred on W_1/W_2. The method rests on the assumption that words entering into the same syntagmatic relation with other words are perceived as semantically similar. The method has a potential for capturing word similarities grounded on contextual effects of the sort sketched out above, although it may often happen that, given two instances of the same word W in a text and their corresponding context windows, very few token words are found in both windows. Strategies to alleviate this sparse data problem have been described for word sense disambiguation (e.g. Schütze 1992): they define the context no longer in terms of the immediate neighbouring words, but rather as the set of words that neighbouring words normally consort with. An interesting issue here is whether "context cascades" of this sort are still constrained enough to be able to capture effects of context-sensitive similarity. The amount of data that this method requires is also an issue.

Be that as it may, it is still to be shown conclusively that any of the NLP tasks listed at the outset really requires such a fine grained measure of context-sensitive semantic similarity. In this paper, we contend that an ideal lexical resource aimed at being used as a yardstick for measuring word sense similarity

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

at the level of granularity required by most NLP applications should strive to complement the lexico-semantic knowledge typically embedded in a WordNet-like resource with distributional evidence of some kind. This is argued on grounds that: i) contextual factors play an important role in assessing the semantic similarity between words and ii) this is what most applications require. Both points will be dealt with in some detail in the context of the problem of classifying the typical complements lexically selected by a given verb sense. We will show that verbs' selectional preferences cannot always be neatly expressed in terms of taxonomy nodes/classes, but rather cut across the taxonomy in a seemingly erratic way, straggling for several relatively unrelated nodes. Close examination of real data shows that different verb senses select different classes of complements according to different dimensions of semantic similarity, to such an extent that it soon becomes impossible to provide an effective account of these dimensions independently of the verb sense in question.

2. Taxonomy-based semantic similarity: general background

Different methods have been put forward in the literature to assess semantic similarity in relation to a hierarchically structured lexical resource such as WordNet. In most of them (see among others Rada et al. 1989 and Lee et al. 1993), assessment of semantic similarity is carried out on the basis of hyperonymy (IS-A) links. More concretely, semantic similarity is evaluated by measuring the distance between the taxonomical nodes corresponding to the items being compared: the shorter the path from one node to another, the more similar the corresponding items. Given multiple paths, the shortest path is taken as the one involving the stronger similarity.

A number of criticisms have been levelled at this approach. Some scholars pointed out that IS-A links are simply not sufficient. Nagao (1992), for instance, uses both hyperonymy and synonymy links to compute semantic similarity, and assigns higher similarity scores to synonymy relationships. Other scholars have attempted to further widen the range of relationships on the basis of which semantic similarity is computed; see, among others, Niremburg et al. (1993) who also use morphological information and antonyms.

A more technical problem faced by the path-length similarity method has to do with the underlying assumption that links in a taxonomy represent uniform distances between nodes. As often pointed out, this is not always the case: in real taxonomies, the "distance" covered by individual taxonomic links is variable, since certain sub-taxonomies can be much denser than others. To overcome the problem of varying link distances, Agirre and Rigau (1996) propose a semantic similarity measure (referred to as "conceptual density") which is sensitive to i) the length of the path, ii) the depth of the nodes in the hierarchy (deeper nodes are ranked closer) and iii) the density of nodes in the sub-hierarchies (concepts involved in a denser subhierarchy are ranked closer than those in a more sparse region). In a similar vein, Resnik (1995) defines a taxonomic similarity measure which dispenses with the path length approach and is based on the notion of information content. Under his view, semantic similarity between two words is represented by the $-\log P(C)$ value of the most informative concept C subsuming both words in a semantic taxonomy, where P(C) is a maximum likelyhood estimate of C's probability of occurrence in a reference corpus.

Despite their differences, all these methods address the issue of how lexico-semantic hierarchies like WordNet should best be exploited, but do not question their suitability for measuring word semantic proximity. This issue will be dealt with in some detail in the following section.

3. Taxonomy-based semantic similarity at work: an illustrative example

In this section, the problem is tackled of how and to what extent a WordNet-like lexical resource can provide the information needed to assess semantic similarity of words in context, in connection with the task of of semantically characterising the class of typical collocates of a given verb sense. In section 3.1 a taxonomy-based account of selectional preferences of different senses of the same verb is illustrated. This is complemented with a comparative study of intersecting sets of typical collocates of different verb senses (section 3.2).

3.1 A taxonomy-based account of selectional preferences of verbs

This section illustrates the modelling of the selectional preferences of different senses of a verb according to a taxonomy-based view. To exemplify, we consider here the different senses of the Italian verb *accendere* together with the sets of their typical object collocates. These typical objects are projected onto a semantic hierarchy to evaluate whether and to what extent the verb's selectional preferences are captured through taxonomical generalizations of some kind.

According to the Collins Italian-English dictionary (1985), the Italian verb *accendere* has, in its transitive reading, the following four senses, each accompanied by an illustrative set of its typical objects:

1) *light* when it takes as a direct object nouns like *fiammifero, candela, sigaretta, camino* (respectively, 'match, candle, cigarette, fireplace')
2) *turn on, switch on*, when the object is a device such as *radio, luce, lampada, gas, motore* (respectively, 'radio, light, lamp, gas cooker, engine')

3) *raise*, if the object is some kind of feeling such as *speranza*, *desiderio* ('hope, desire')
4) *open*, if the object is a bank-related entity such as *conto*, *debito*, *ipoteca* 'bank account, debt, mortgage'

Given the source of lexical information considered here, each sense is characterised in terms of its appropriate English translation equivalent. The tree-like structure reported below illustrates the result of projecting Collins' typical object collocates of each set onto WordNet.[1]

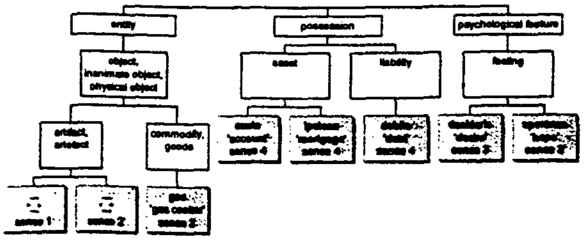

Figure 1 Objects of *accendere*: semantic hierarchy

Shadowed boxes (typically but not necessarily tree leaves) represent the actually occurring collocates, which are accompanied by an indication of the sense of *accendere* with which they are associated in the dictionary. Dotted lines in the tree show that the link between the connected nodes is not direct, i.e. that the taxonomical path includes intermediate nodes.

The first thing to note in this context is that collocates of different senses exhibit a different propensity to cluster together in the semantic hierarchy. The selectional preferences of senses 3 ('raise') and 4 ('open') nicely fall into distinct branches of the taxonomy. The class of typical objects of sense 3 can appropriately be described as a <feeling>, while <possession> being a suitable hyperonym of all and only objects of sense 4. Yet, the same taxonomy fails to part the selectional preferences of sense 1 ('light') from those of sense 2 ('switch on'). In the latter case, object collocates of both senses are categorised as an <artifact>, a notion which is far too general to tell the collocates of sense 1 of *accendere* from those of its sense 2, as illustrated by the internal structure of the sub-hierarchy of artifact objects of *accendere* diagrammed in Figure 2 below. For senses 1 and 2 of *accendere*, the clustering of nodes in the subtaxonomy of 'artifacts' does not help to identify the semantic "glue" that keeps together the object collocates for each relevant sense.

[1] Although we are working on Italian examples, we will use hereafter, for illustrative purposes, WordNet 1.5 as a reference taxonomical resource, due to its completeness, the Italian WordNet being still under development in the framework of the EuroWordNet project (LE2-4003). This decision is not arbitrary, since, for the words considered in this paper, the Italian WordNet shows a similar taxonomical organization as WordNet 1.5.

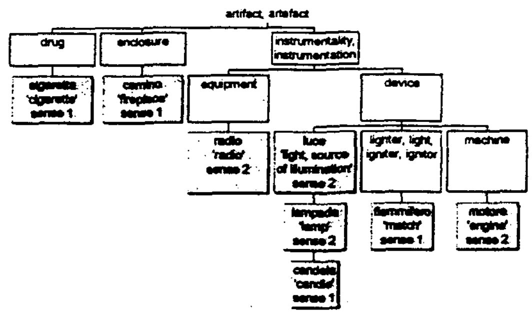

Figure 2 Hierarchy of 'artifact' objects of *accendere*

Sense 1 of *accendere* selects for artifact objects which can burn; sense 2 basically selects for devices which are activated through making electric contact.

The problem here is not simply that it is impossible to identify one single upper node covering - say - all and only burning artifacts as opposed to devices making electric contact. The classical assumption that one semantic class should be made to contain all and only the collocates of one sense is clearly too strong in this context, if it is a workable cognitive hypothesis at all. One could nonetheless fall back to the weaker assumption that a class of collocates be expressed in terms of a disjunction of the taxonomy's nodes/subclasses, provided that each such node/subclass defines a proper subset of the typical objects of the verb sense in question. In fact, our diagram above shows that even this weaker characterization is not viable in all cases. Consider the taxonomy chain formed by *luce-lampada-candela* corresponding to the class of objects having to do with <light, source of illumination>. Whereas *luce* 'light' and its hyponym *lampada* 'lamp' both point to the 'switch on' sense (sense 2), *candela* 'candle' (the terminal node of this chain) is associated with the 'light' sense (sense 1), due to its being a typically burning object. Here the same taxonomy chain includes objects related to different senses of the verb. This is tantamount to saying that the dimension of semantic similarity captured through the taxonomical structure is not appropriate but rather misleading if one wants to unambiguously characterise the different senses of the verb through their selectional preferences. The property of burning, on which the preference is based, cannot possibly be percolated from higher to lower nodes through the taxonomy chain. Rather, it represents a property peculiar of some nodes only, either intermediate or terminal ones. Hence, given the taxonomy illustrated above, one can do little more than disjunctively listing all nodes corresponding to the collocates in question, with the further stipulation that the property of being a collocate does not necessarily percolate further down in the taxonomy chain. This is fine, but it boils down to saying that the taxonomy in question can do very little to generalize over the selectional preference classes.

To sum up, this simple example shows that taxonomy-based semantic similarity is not always sufficient to justify the belonging of a given lexical item to a specific selectional preference class. Very granular distinctions may be needed to characterise any such class. Moreover, some of the distinctions required are orthogonal to the distinctions conveyed by a taxonomical organisation of the lexicon.

3.2 Comparing overlapping selectional preferences of different verbs

So far, we focussed on the difficulty of neatly characterising verb selectional preferences in terms of taxonomical classes. It turns out that the semantic glue pasting together the object collocates of senses 1 and 2 of the verb *accendere* is given by distinctions which are not directly reflected in the semantic taxonomy. Taxonomical relationships seem to capture only some of the various dimensions on which semantic similarity is grounded. This is not accidental, we believe, since taxonomical dimensions are typically defined i) independently of context, and ii) once and for all. It is thus not surprising that they may fail, in some cases, to reflect the similarity dimension appropriate in a specific context. In this section, this issue is explored in more detail by comparing the selectional preferences of different verbs exhibiting a non empty intersection of the sets of their typical collocates.

Among the typical collocates of sense 1 of *accendere* 'light' there is *sigaretta* 'cigarette' which, in WordNet 1.5, is the terminal node of the following taxonomical path:

sigaretta 'cigarette'
=> roll of tobacco
=> tobacco, baccy
=> narcotic
=> drug
=> artifact, artefact
=> object, inanimate object, physical object
=> entity

Let us look now at some of the typical verbs with which *sigaretta* occurs, together with other possible collocates of these verbs, as they are attested in the Collins Italian-English Dictionary (1985), in both example sentences and the semantic indicators field. In these examples, the sequences "/S" (short for "Subject") and "/O" (short for "Object") specify the grammatical relation of the noun relative to the verb:

- ACCENDERE$0_1/V {SIGARETTA/O CAMINO/O CANDELA/O FIAMMIFERO/O}
  light {cigarette/O, fireplace/O, candle/O, match/O}
- ARROTOLARE$0_1/V {SIGARETTA/O CARTA/O STOFFA/O}
  roll up {cigarette/O paper/O fabric/O}
- FUMARE$0_1/V {SIGARETTA/O PIPA/O}
  smoke {cigarette/O pipe/O}
- OFFRIRE$0_1/V {SIGARETTA/O AIUTO/O LAVORO/O MERCE/O PREGHIERA/O}
  offer {cigarette/O help/O job/O goods/O prayer/O}
- RIACCENDERE$0_1/V {SIGARETTA/O FUOCO/O GAS/O INTERESSE/O LUCE/O RADIO/O SENTIMENTO/O}
  light/switch on/revive {cigarette/O fire/O gas/O interest/O light/O radio/O feeling/O}
- SPEGNERE$0_1/V {SIGARETTA/O APPARECCHIO/O DEBITO/O FUOCO/O GAS/O LUCE/O PASSIONE/O SUONO/O}
  extinguish/switch off/stifle/muffle {cigarette/O device/O debt/O fire/O gas/O light/O passion/O sound/O}
- SPEGNERSI$0_2/V {SIGARETTA/S APPARECCHIO/S FUOCO/S LUCE/S PASSIONE/S RICORDO/S SUONO/S}
  be extinguished/stop/fade away {cigarette/S device/S fire/S light/S passion/S memory/S sound/S}

Careful consideration of these examples shows that different types of semantic glue are at work in different cases. With the verb *accendere* (sense 1) the glue is, as we saw, the property of burning. A similar analogy is at work in the case of *riaccendere*, *spegnere* and *spegnersi*, with the main difference that this case also includes figurative usages. As to the verb *arrotolare*, the semantic similarity of its object collocates is grounded on their being made of material whose texture makes them rollable. The relevant similarity which links pipes and cigarettes relative to the context of *fumare* rather hinges on their telic role, their both being typically smoked objects. Finally, the collocational set of *offrire* includes words denoting typical human needs and/or desires ranging from cigarettes and goods to more abstract things such as help and prayers.

These examples confirm the difficulty of assessing the semantic similarity of words when they are considered outside their actual contexts of use, difficulty which already emerged in relation to a characterization of the selectional preferences of the verb *accendere*. By projecting these collocational sets onto WordNet, appropriate generalisations can hardly be found. A general semantic class subsuming some or all members of each set may exist, but often it is not specific enough to avoid undesired intersection of classes, as in the case of senses 1 and 2 of *accendere*. On the other hand, semantic features such as "lightability", "enjoyability", "smokability" or "rollability" seem to be at work here: they strike us as hardly amenable to a global consistent taxonomical rendering.

3.3 Implications

In the previous sections, we discussed whether and to what extent taxonomical relationships as actually implemented in WordNet-like lexical resources can be used to measure the semantic similarity of typical collocates associated with a given verb sense. We showed that one can hardly find a unique taxonomy node subsuming all and only the collocates bearing the same grammatical relation to a given verb sense. A weaker but more realistic hypothesis was also considered, namely that a class of verb collocates be expressed in terms of a disjunction of taxonomy's nodes, provided that each such node defines a proper subset of the typical collocates of the verb sense in question. It turned out that even this weaker characterization of selectional preferences is not always viable since it is often the case that selectional preference information is not disjunctively distributed over taxonomy nodes. When this is the case, a taxonomy provides virtually no means of generalising over the set of typical collocates of a given verb sense.

In our view of things, such an inadequacy of taxonomical information cannot be got around by letting finer grained distinctions slip in the semantic type model. Rather, it bears upon one inherent property of most taxonomies as they are currently built up: monodimensionality. In fact, taxonomies are often anchored to a fixed classificatory dimension (e.g. perceptual features as opposed to functional ones). By contrast, real data suggest that different verb senses select different classes of complements according to different dimensions of semantic similarity. This is the reason why taxonomies do not always capture locally salient common features, which are needed to appropriately account for the semantic similarity of verb complements.

Our examples showed that multidimensional classifications are indeed required to dynamically capture locally salient features. Although in WorNet 1.6 provision is made for concepts to be cross-classified with respect to different dimensions, it is not clear how many and what dimensions should be added to the original WordNet structure to comply with real NLP application requirements. These considerations are, in our view, compelling enough to prompt the investigation of different and more workable ways to complement the taxonomical structure of WordNet-like resources. A simple but effective source of knowledge which can nicely complement WordNet for capturing locally salient semantic similarity is represented by distributional information about words, under the assumption that words which bear the same syntactic relation to the same word sense form a somehow semantically coherent class.

In the following section, we illustrate this point by describing a measure of semantic similarity based on distributional evidence and we show how helpful this is in capturing locally salient semantic similarity.

4. Distributionally-based semantic similarity

A semantic similarity measure computed on the basis of distributional evidence is at work in SENSE, an example-based word sense disambiguation (WSD) system carrying out the task on the basis of a representative set of typical patterns of use (Federici et al. 1997). In particular, SENSE presupposes prior availability of verb-noun pairs where the contextually relevant sense of the verb token is assigned. At the same time, the accompanying noun is provided with its grammatical function. This set of verb-noun pairs constitutes the knowledge base of examples (or example base for short) on the basis of which SENSE is able to draw its inferences.

Given an Input Pair IP to be disambiguated where the grammatical relation of the noun relative to the verb is specified, SENSE searches its example base looking for the set of examples which are most similar to IP. If an identical pair is found in the example base, then the usual assumption is made that the verb token in IP is used in the same reading of the verb in the known example.[2] The key notion used by SENSE to compute similarity between non identical pairs is proportional analogy. To illustrate, if the verb sense in the pair *accendere-pipa*/O 'light-pipe' has to be inferred, this can be done through the following proportion, involving three disambiguated verb-object pairs attested in the example base plus the input pair *accendere-pipa*/O as the fourth term:

fumare_1-sigaretta/O	:	*fumare_1-pipa*/O	=	*accendere_1-sigaretta*/O	:	*accendere_?-pipa*/O
'smoke-cigarette/O'	:	'smoke-pipe/O'	=	'light-cigarette/O'	:	'light-pipe/O'

Intuitively, the proportion says that the sense of *accendere* in *accendere-pipa* 'light-pipe' is likely to be the same as in *accendere-sigaretta* 'light-cigarette' since both *pipa* and *sigaretta* can typically be smoked, or - in more linguistic terms - since they are both typical objects of sense 1 of *fumare* 'smoke' (*fumare_1*).

It is important to point out here that this inferential strategy is "local" in two senses: i) relative to the example base, and ii) relative to the input pair. First, it neither presupposes nor relies on a preliminary classification of all known examples. In this respect, the system simply memorizes all examples, with no attempt to generalize over them in any optimal global way. Generalizations are only made to interpret new unknown evidence. Hence, the resulting classification does not reflect general properties of the example base as such, but only associations which are triggered by the specific input pair in question. In this sense the hypothesis search space is constructed on the fly, every time the system is confronted with a new unknown pair.

The second notion of "locality" we intend to emphasize here is related to the issue of what constitutes a relevant analogy, given the input pair IP considered. The similarity between an IP and some known examples is not simply based on a a-priori

[2] In fact, SENSE is also able to go beyond the evidence provided by an attested example as illustrated in Federici et al. 1997.

global similarity of some of its constituent elements (i.e. the verb and the noun), which, as we just saw, is not available. An analogical proportion enforces a much more constraining relation. The interpretation 'light-pipe' of *accendere-pipa* is not simply based on the piecemeal analogy with *accendere-sigaretta* (where *accendere* is found in common, and pipe and cigarette are sufficiently similar). The conclusive element of the analogy is that both *fumare* and *accendere* in their respective senses of 'light' and 'smoke' are systematically related in the example base through a set of shared objects, and that *pipa* occurs with *fumare* in the required sense. This is exactly what the proportion is able to capture.

We contend that, for the notion of context-sensitive word sense similarity to adequately be modeled, both notions of locality play an important role.

The example base used so far for testing the effectiveness of the distributionally-based semantic similarity measure for WSD purposes was automatically acquired from both semantic indicators and example sentences of the Collins Italian-English Dictionary (Montemagni 1995). Each acquired verb-noun pair can thus be said to represent a typical pattern of use of a given sense of a verb. The choice of a bilingual dictionary was also motivated by the practical interest that the resulting sense subdivisions have for purposes of Machine Translation.

The derived example base contains 8,153 verb-noun pairs (either verb-subject or verb-object patterns) which exemplify 3,359 different verb senses. All pairs are acquired from verb entries, and thus provide sense information only about the verb; each accompanying noun, if polysemous, is not disambiguated. On average, a verb sense is illustrated through 2.42 patterns. Senses which are attested in ten or more patterns are a negligible part of the training set, whereas most verbs are illustrated through a number of patterns ranging between 2 and 5. Finally, a considerable group of verb senses is attested only once. Note that the latter circumstance does not stop SENSE from recognising hapax senses in novel unknown contexts.

SENSE performance was tested on a corpus of 150 IPs randomly extracted from unrestricted texts. Since the test was intended to evaluate the reliability of distributionally-based inferences, the test corpus did not contain any pattern already present in the example base. Only verbs were disambiguated. The results of this experiment are reported in the table below:

	Overall	Polysemous
RECALL	79.3%	66.3%
PRECISION	89.9%	80.4%

Figures in the first column refer to both polysemous and monosemic verbs. In the second column, recall and precision are relative to polysemous verbs only. These figures are very significant if one considers i) the comparatively small size of the lexical database used for training, ii) the distribution of patterns per verb sense, and iii) the fact that only some of its attested words (namely verbs) are semantically disambiguated.

The results reported above were computed on the basis of distributional evidence only. On closer analysis, it turned out that some of the input contexts which were left ambiguous by SENSE could have been successfully disambiguated if also taxonomical information was taken into account. Consider the following three cases:

	verb	sense	object	
input	*abbattere*	?	*pianta*	hyponym:
context	'cut down'		'plant'	*albero*
known	*abbattere*	1	*albero*	
example	'cut down'		'tree'	
input	*abbassare*	?	*capo*	synonym:
context	'hang'		'head'	*testa*
known	*abbassare*	1	*testa*	
example	'hang'		'head'	
input	*accarezzare*	?	*barba*	hyperonym:
context	'stroke'		'beard'	*pelo* 'hair'
known	*accarezzare*	1	*capello*	hyperonym:
example	'stroke'		'hair'	*pelo* 'hair'

In the first case, the object in the target context is the WordNet hyperonym of the object in the known example, as shown in the rightmost column of the table. In the second case, the objects of both input and known pairs are synonyms. Finally, the last case illustrates a typical instance of hyperonym sharing. This indicates that distributionally-based and taxonomy-based inferences can nicely be complemented. In practice, this can be done in more than one way. In some experiments of syntactic disambiguation (subject/object assignment in Italian, Montemagni 1995, Montemagni et al. 1996), we tried to combine both taxonomical and distributional measures in such a way that the system relied on taxonomical information first, to turn to distributional evidence only when the first step was not conclusive. This strategy, however, did not seem to be successful, as the system was frequently led astray by irrelevant similarities. Our experience seems to suggest that a more promising way to integrate distributionally-based and taxonomy-based information is arguably to use distributional evidence first, so as to exploit the context-sensitivity (or locality) of proportional analogy as a filter of irrelevant similarities. Taxonomical information is to be relied on only at a second stage, as a fall back solution to outstanding ambiguities.

5. Conclusions

Semantic similarity is not simply a relation between two words in isolation, but rather a relation between two words in their context. This context-sensitive view of semantic similarity makes its identification more problematic. In principle, semantic similarity of words can be captured in a number of different ways, ranging from their taxonomical relationships to their actual distribution in a corpus. It would be very difficult to argue that one such a way is more plausible than another; nonetheless, it should be observed that their practical utility in well-known interesting NLP applications can vary considerably.

We noted that taxonomy-based measures of semantic similarity are to an extent inadequate, as they capture only some of the classificatory dimensions which play a relevant role in NLP applications. We showed that relevant similarities need to be grounded on the specific context to be processed (e.g. disambiguated, retrieved or summarised) and that different contexts call for different classificatory dimensions. Distributional evidence can be used to model this sort of context-sensitive multidimensional classification, so as to induce semantic associations between words that nonetheless belong to different places in a taxonomy. We also showed that distributionally-based semantic similarity has a considerable impact on crucial NLP tasks such as word sense disambiguation. All this provides evidence that WordNet-like lexical resources should strive to integrate taxonomical and distributional information, by combining both paradigmatic and syntagmatic dimensions.

As already mentioned, WordNet has a potential for doing that, through extended implementation of so-called pointers from nouns to verbs and from verbs to nouns, to represent functions, typical semantic preferences etc. Within the EuroWordNet project (LE2-4003), some steps in this direction have already been taken in developing multilingual WordNets for Dutch, Italian and Spanish. Among the additions to the original set of relations borrowed from WordNet 1.5, syntagmatic relations feature prominently: e.g., one finds verb-to-noun relations denoting the typical entities involved in a given event, or noun-to-verb relations referring to the typical events in which a given entity play a role (Alonge et al. forthcoming).

This certainly provides the information needed to capture context-sensitive semantic similarities. We also showed that local inferential engines such as SENSE can demonstrably tap this type of information with the degree of flexibility, noise-tolerance and input-relevance required, among others, by WSD.

References

Alonge A., N. Calzolari, P. Vossen, L. Bloksma, I. Castellon, A. Marti, W. Peters, forthcoming, 'The Linguistic Design of the EuroWordNet Database', in *Computers and the Humanities: Special Issue on EuroWordNet.*

Agirre E., G. Rigau, 1996, 'Word Sense Disambiguation using Conceptual Density', in *Proceedings of COLING-96*, Copenhagen, pp. 16-22.

Brown P.F., S.A.D. Pietra, V.J.D. Pietra, R.L. Mercer, 1991, 'Word Sense Disambiguation Using Stastical Methods', in *Proceedings of ACL-1991*, pp. 264-270.

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

Federici S., S. Montemagni, V. Pirrelli, 1997, 'Inferring semantic similarity from Distributional Evidence: an Analogy-based Approach to Word Sense Disambiguation', in *Proceedings of the ACL/EACL Workshop "Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications"*, Madrid, Spain.

Gale W.A., K.W. Church, D. Jarowsky, 1992, 'A Method for Disambiguationg Word Senses in Large Corpora', *Computers and the Humanities.*

Goldstone R.L., Medin D.L., Halberstadt J., 1997, 'Similarity in context', *Memory and Cognition*, 25(2), pp. 237-255.

Lee J.H., M.H. Kim, Y.J. Lee, 1993, 'Information Retrieval based on conceptual distance in IS-A hierarchies', *Journal of Documentation*, 49(2), June 1993, pp. 188-207.

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

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

Nagao M., 1992, 'Some Rationales and Methodologies for Example-Based Approach', in *Proceedings of "International Workshop on Fundamental Research for the Future Generation of Natural Language Processing"*, 30-31 July 1992, Manchester, pp. 82-94.

Nirenburg S., C. Domashnev, D.I. Grannes, 1993, 'Two Approaches to Matching in Example-Based Machine Translation', in *Proceedings of TMI-93*, pp. 47-57.

Pereira F., N. Tishby, 1992, 'Distributional Similarity, Phase Transitions and Hierarchical Clustering', Working Notes, Fall Symposium Series, AAAI, pp. 108-112.

Rada R., M. Hafedh, E. Bicknell, M. Blettner, 1989, 'Development and application of a metric on semantic nets', IEEE Transactions on System, Man, and Cybernetics, 19(1), pp. 17-30.

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

Schütze, H., 1992, 'Word Sense Disambiguation with Sublexical Representations', Workshop Notes, Statistically-based NLP Techniques, AAAI, 109-113.