Natural Language Engineering **5** (2): 207–218. Printed in the United Kingdom

© 1999 Cambridge University Press

SENSE: an analogy-based Word Sense Disambiguation system

STEFANO FEDERICI, SIMONETTA MONTEMAGNI
and VITO PIRRELLI
Istituto di Linguistica Computazionale, CNR, Pisa, Italy

(*Received 3 November 1998; revised 15 April 1999*)

Abstract

The paper describes SENSE, a word sense disambiguation system which makes use of multidimensional analogy-based proportions to infer the most likely sense of a word given its context. Architecture and functioning of the system are illustrated in detail. Results of different experimental settings are given, showing that the system, in spite its conservative bias, successfully copes with the problem of training data sparseness.

1 Introduction

Sense disambiguation of a word occurrence in a given context (hereafter WSD) requires appeal to a wide typology of cues, ranging from syntactic subcategorization to lexico-semantic information and subject domain. Such a motley variety of linguistic factors characterises WSD as a linguistically multidimensional task, and makes challenging demands on the architecture of automatic WSD systems. The system described here takes the issue of linguistic multidimensionality seriously, although only a comparatively small subset of the relevant dimensions will be the focus of this paper. We take pains to show how various types of linguistic information can in principle be integrated as multiple constraints on the selection of the most likely reading of a word in context through the use of multidimensional analogy-based proportions. We will also show the variation of performance of our system as a function of the varying complexity of the problem space, in terms of the number of its defining dimensions. Most emphasis will, nonetheless, be laid on the use of lexico-semantic information through proportionally-based similarity measures.

2 Basic requirements of a WSD system

WSD systems typically rely on (1) predefined sense distinctions, (2) a formal characterization of the notion of context, and (3) means to assign the most likely sense to a given word on the basis of its context.

The idea of modelling the space of word meanings of a given entry as a list of sense

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

tags, however psycholinguistically hard to justify (Kelly and Stone 1975), is taken for granted in almost all WSD work to date, and will not be discussed here. As for (2), the last ten years have witnessed a prominent shift of focus from a linguistically loose notion of context (typically, an *n*-word text window centred on the polysemous word to be classified), to a considerably more local one (e.g. the subject/object of the verb token to disambiguate). Point (3) covers a battery of different techniques for assigning a sense tag to a word W in a target context T, ranging from statistically-based approaches, to decision trees and example-based approaches (see Ide and Veronis (1998) for a recent survey). While the first two approaches strive for a global characterisation of the polysemy space of each word (in terms of independent classifiers consisting of either a matrix of token frequency distributions or a decision tree), example-based approaches require virtually no learning, and rely on as many local classifiers as sense-tagged instances of corpus contexts of W. Finally, systems may or may not rely on some independent repositories of lexico-semantic knowledge such as thesauri or taxonomies.

In spite of this considerable range of variation, all WSD systems make explicit or implicit use of a shared set of assumptions concerning the notion of inter-contextual similarity as a basis for WSD. It is to this notion that we turn now.

3 On inter-contextual similarity

The following assumption plays a prominent role in most WSD systems:

(1) Two occurrences of the same word have an identical meaning if they have similar local contexts.

As pointed out by Lin (1997), assumption **(1)** requires that, for each polysemous word, a distinct classifier be built. There is no way to use the classifier learned for one word to learn the classifier of a different word. This has serious repercussions on the amount of training required. Moreover, systematic regularities concerning, e.g. word meaning shifts or the regular correlation of selectional restrictions with specific classes of predicates are overlooked completely in **(1)**, whatever the algorithm used for sense tagging. In short, assumption **(1)** seems not to take advantage of linguistic relations among data, and takes the similarity of two contexts at face value.

This is, in our opinion, linguistically haphazard and risky. Inter-contextual similarity is not a sufficient condition to ensure sense invariance. As shown by Montemagni and Pirrelli (1998), there is no guarantee that there always exists a suitable 'tree cut' or class of neighbouring nodes (Abe and Li 1996), in a taxonomy like WordNet, capable of discriminating all and only the true collocates of a verb sense. Words which are extremely close in a taxonomy can still form part of the contexts of two very different senses of W. For example, neighbouring nodes in the Italian (Euro)WordNet (Vossen, Ide and Greenstein 1998), such as *candela* 'candle' and *lampada* 'lamp', are objects of two very distinct senses of the Italian verb *accendere*, 'light' and 'switch on' respectively. Conversely, the collocates of a given word sense can well include nodes which are far removed in a taxonomy: see, for instance, *candela* 'candle', *sigaretta* 'cigarette' and *camino* 'fireplace' as objects of *accendere*

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

'light'. Lin (1997) proposes the following alternative strategy as a way to alleviate some of the limitations of **(1)**:

(2) Two different words are likely to have similar meanings if they occur in identical contexts.

On the basis of **(2)**, the sense-tagged context of one word is used to disambiguate a different word, on condition of context identity. Assumption **(2)**, however, only shifts the use of a taxonomy from assessment of inter-contextual similarity to sense assignment. As such, **(2)** is still prone to the same principled fallacy of **(1)**: given a target context T and an ambiguous word W, there is no guarantee that the sense of a different word replacing W in T be a taxonomic neighbour of an appropriate sense of W. This problem compounds with the requirement of context identity to make sparse data still an issue. Finally, **(2)**, like **(1)**, does not seem to be able to exploit systematic relations among sense-tagged contexts. We suggest that **(1)** and **(2)** be replaced by the following:

(3) Given two occurrences of the same word, they will have identical meanings if and only if their local contexts are shown to preserve the sense of at least two occurrences of another word.

According to **(3)**, two contexts are not assumed to be sense preserving on the basis of an existing, and possibly occasional, inter-contextual similarity. It is a systematic relationship of sense invariance that establishes a similarity between two contexts. As a result, two contexts are judged similar or not always in relation to a target W. Note further that two contexts can be sense preserving but syntactically very different. For example, in the causative-inchoative alternation (e.g. *John broke the vase*, *the vase broke*) objects of the transitive reading of *break* are systematically found as subjects of the intransitive reading of the same verb. This means that the sense preserving objects of the transitive reading of *break* (e.g. *John broke the glass/the vase*) remain sense preserving as subjects of its intransitive reading (e.g. *the glass/the vase broke*). While **(3)** is able to extrapolate sense invariance from a given construction to a syntactically different one, this cannot be done through either **(1)** or **(2)** above, under a literal interpretation of context similarity/identity. Inter-contextual sense invariance realised through syntactically-asymmetric constructions is not restricted to verbs undergoing diathesis alternations but also includes verbs with overlapping selectional preferences relative to different syntactic functions.

To sum up, **(3)** has the potential of making the most of sparse data in a linguistically controlled way; this also presents the nonnegligible advantage of grounding similarity on sense invariance, rather than using the former to infer the latter. In the remainder of the paper we describe in some detail how assumption **(3)** can be implemented and put to practical use for WSD.

4 SENSE: a proportionally-based WSD system

SENSE (Self-Expanding linguistic kNowledge-base for Sense Elicitation) is a specialised version of a generalised language-learning system (Federici and Pirrelli 1994;

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

Table 1. *The WCP fumare-sigaretta*

'smoke–cigarette'	
PRED:	*fumare*
S_PRED:	2
SYNT_FUNCT:	obj
ARG:	*sigaretta*
S_ARG:	1

Federici, Montemagni and Pirrelli, 1996) tailored for WSD (Federici, Montemagni and Pirrelli 1997). Broadly speaking, it belongs to the family of example-based WSD systems in that it assigns a sense tag to a target word W in context T on the basis of the analogy of T to an already sense-tagged context of W. SENSE uses a very local notion of context, modelled as a word co-occurrence pattern (WCP). WCPs are stored in an Example Base (EB) and used directly to assign sense tags to a word in a target context.

4.1 Word co-occurrence patterns and analogical proportions

WCPs are complex linguistic objects, modelled through sets of attribute-value pairs as exemplified in Table 1. The top line in the table is a mnemonic identifier of the whole feature structure. The verb element is described in terms of its lemma (PRED) and sense (S_PRED). The lemma and the sense of the noun element are specified by the features ARG and S_ARG, respectively. The feature SYNT_FUNCT gives the syntactic function of the noun relative to the verb. These attributes pertain to different levels of linguistic description (syntactic and lexico-semantic), representative, but by no means exhaustive, of the factors which may have a bearing on WSD. Moreover, the list is non minimal, since WSD can be carried out effectively by leaving some of these dimensions unspecified (see section 5).

WCPs show a considerable degree of a systematic mutual relatedness and can thus be used to establish multidimensional analogical proportions (Pirrelli and Yvon 1999) such as that exemplified in Table 2, where each term in the proportion is assigned a distinct column, and represents a WCP. WCP-defining attributes are arranged along distinct rows. Terms are numbered from t_1 to t_4. Given an attribute, we call v_i the value of that attribute relative to t_i. For each attribute, its values are said to be proportionally related if at least one of the following conditions holds:

(1) $(v_1 = v_2)$ and $(v_3 = v_4)$;
(2) $(v_1 = v_3)$ and $(v_2 = v_4)$.

The four terms t_1–t_4 are said to be proportionally related if, for each of their defining attributes, the corresponding values v_1–v_4 are proportionally related. These conditions are met in proportion 1 above as follows: predicate-related features (i.e. PRED and S_PRED) comply with condition 1), while argument-related features (i.e. ARG and S_ARG) meet condition 2). Note further that the values of SYNT_FUNCT fullfil

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

Table 2. *Analogical proportion 1*

'smoke–cigarette'	:	'smoke–pipe'	=	'light–cigarette'	:	'light–pipe'
t_1		t_2		t_3		t_4
PRED: *fumare*		PRED: *fumare*		PRED: *accendere*		PRED: *accendere*
S_PRED: 2		S_PRED: 2		S_PRED: 1		S_PRED: 1
SYNT_FUNCT: obj		SYNT_FUNCT: obj		SYNT_FUNCT: obj		SYNT_FUNCT: obj
ARG: *sigaretta*		ARG: *pipa*		ARG: *sigaretta*		ARG: *pipa*
S_ARG: 1		S_ARG: 1		S_ARG: 1		S_ARG: 1

Table 3. *Analogical proportion 2*

'show–painting'	:	'show–photo'	=	'hang–painting'	:	'hang–photo'
t_1		t_2		t_3		t_4
PRED: *rappresentare*		PRED: *rappresentare*		PRED: *attaccare*		PRED: *attaccare*
S_PRED: 2		S_PRED: 2		S_PRED: 1		S_PRED: 1
SYNT_FUNCT: subj		SYNT_FUNCT: subj		SYNT_FUNCT: obj		SYNT_FUNCT: obj
ARG: *quadro*		ARG: *fotografia*		ARG: *quadro*		ARG: *fotografia*
S_ARG: 1		S_ARG: 1		S_ARG: 1		S_ARG: 1

conditions (1) and (2) simultaneously, as all terms in the proportion exhibit the same value for the same attribute. What proportion 1 says about its terms can be paraphrased as follows: *sigaretta* 'cigarette' and *pipa* 'pipe' are semantically-related nouns since they are S_PRED preserving across syntactically-symmetric contexts of two different verbs (*fumare* 'smoke' and *accendere* 'light'); *fumare* and *accendere*, in their turn, are semantically-related, since they share a (sub)set of object collocates.

In proportion 1, nouns stand in the same syntactic relation to verbs. However, as already pointed out in section 3 above, it is often the case that clusters of nouns which function, say, as the object of a given verb can also function as typical subjects of other, somehow related, verbs. This is captured through proportions of verb–noun pairs involving syntactically-asymmetric constructions, as exemplified in Table 3, where the feature SYNT_FUNCT appears to follow the same proportional pattern of predicate-related features (condition 1 above). What is entailed by proportion 2 is the following: *fotografia* 'photo' and *quadro* 'painting' are semantically-related nouns, since they are S_PRED preserving across syntactically-asymmetric contexts of two different verbs (*rappresentare* 'show' and *attaccare* 'hang'); *rappresentare* and *attaccare* are semantically related since the subject collocates of the former and the object collocates of the latter have a non empty intersection.

4.2 Multidimensional analogical proportions for WSD

Multidimensional analogical proportions can include partially specified WCPs like t_4 in Table 4, where the value of S_PRED is unknown. The relations holding among

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

Table 4. *Analogical proportion 2 with underspecified WCP*

'show–painting'	:	'show–photo'	=	'hang–painting'	:	'hang–photo'
t_1		t_2		t_3		t_4
PRED: *rappresentare*		PRED: *rappresentare*		PRED: *attaccare*		PRED: *attaccare*
S_PRED: 2		S_PRED: 2		S_PRED: 1		S_PRED: **?**
SYNT_FUNCT: subj		SYNT_FUNCT: subj		SYNT_FUNCT: obj		SYNT_FUNCT: obj
ARG: *quadro*		ARG: *fotografia*		ARG: *quadro*		ARG: *fotografia*
S_ARG: 1		S_ARG: 1		S_ARG: 1		S_ARG: 1

attributes defining the terms in the proportion can be used to predict the value of S_PRED in t_4 on the basis of one of the following conditions:

(1′) if $(v_1 = v_2)$ then $(v_4 = v_3)$;
(2′) if $(v_1 = v_3)$ then $(v_4 = v_2)$.

By (1′), S_PRED in t_4 is assigned the value '1' by transferring v_3 to the unknown v_4. Hereafter we will refer to the partially specified term entering a proportion as the 'target term'. Furthermore, we call t_1 the 'pivot term'. Note that t_1 – unlike t_2 and t_3 – is the only term in a proportion which may share nothing with the target. Finally, we call a 'transfer term' the term whose value is transferred to the target. Given a target context T containing a word W to be disambiguated, the pivot plays the central role of bridging the gap between the transfer and the target. Thus, the pivot must be specified for different lexical fillers from those in the target.

4.3 Proportionally-based inferences for WSD

Given T and W, sense assignment is carried out through the following steps:

Step 1 If T is already attested in EB as WCP_i, then the sense of W in WCP_i is taken to be the most likely interpretation of W in T; note that this step does not stop SENSE from taking Step 2.

Step 2 If in EB there are any three WCPs to which T is proportionally related as the target term, then the sense of W in T is inferred through conditions $1' - -2'$) in section 4.2 above; all possible solutions are eventually ranked.

Step 3 If Steps 1 and 2 yield no result, no disambiguation is attempted by SENSE.

When more than one sense interpretation of a word W in a target context T is supported, the different sense assignments are ranked according to their likelihood in the specific context. First, SENSE looks for all analogical proportions which contain any three WCPs attested in EB and T as the target term. Each proportion will support a given sense interpretation. The system then ranks all proportionally-supported sense interpretations as a function of (i) number of proportions supporting that sense interpretation, (ii) semantic entropy of the lexical fillers of the pivot terms in the supporting proportions. Semantic entropy is calculated as $\log_2(\text{freq}(w_i))$, according to Melamed's (1997) definition, where 'freq' is interpreted for our purposes

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

as the type frequency of w_i in EB. Finally, W in T is assigned that sense interpretation which comes first in the ranked list of candidate senses. This reflects the intuition that a word which keeps company with a smaller set of different collocates is semantically heavier, and that proportions containing semantically heavier words in the pivot term are more significant than proportions with semantically lighter words.

As stated in Step 1, the inferential routine of SENSE can go beyond the evidence attested in EB. This is often useful, especially when sense distinctions are fine grained and data in EB are sparse. Identification of alternative sense assignments, although with lower ranking, is reasonable, we believe, since WSD is often a matter of suggesting a set of more or less plausible interpretations in context rather than giving 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.

5 SENSE: experimental results

Two experiments (henceforth EXP1 and EXP2) are reported here, focusing on the use of lexico-sematic information for disambiguating verb tokens in context.

Both experiments use an EB of functionally annotated verb–noun patterns automatically extracted from verb entries of mono- and bilingual Italian dictionaries (Montemagni 1995). Despite the well known shortcomings of dictionary data, we decided to use machine readable dictionaries, since they provide ready-made information about word senses and their typical contexts of use.

Number and typology of WCP defining features (dimensions) vary in the two experiments. In EXP1, only verbs are sense-tagged; in EXP2, no sense information is given for both nouns and verbs, with the only exception of WCPs containing the verb to be disambiguated, which is sense-tagged. In both cases, sense distinctions are taken from the dictionary selected as a reference resource.

5.1 Experiment one

For EXP1 we use an EB of 8153 distinct verb–noun patterns (2488 verb–subject, 5665 verb–object) automatically extracted from all verb entries of a bilingual Italian–English dictionary (Collins 1985). In these patterns only verbs are sense-tagged. The patterns exemplify 3359 different verb senses, each exemplified, on average, through 2.42 patterns. A considerable group of verb senses (38%) is attested only once. Note that hapax senses can still be assigned by SENSE to unseen contexts. The crucial point here is that WCPs containing hapax senses cannot function as a pivot in a proportion, since a WCP must be attested at least twice to be able to bridge the target and the transfer terms.

The test corpus consists of 150 verb–subject/object WCPs randomly extracted from unrestricted texts. An extracted WCP is discarded if (a) it was already extracted as a target, (b) it was already in EB, or (c) it contained a word unattested in EB. The 150 contexts are then automatically converted into functionally annotated verb–noun pairs, with the verb as the target word.

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

Table 5. *Results of EXP1*

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

The results of this experiment are reported in Table 5. Figures in the first column concern both polysemous and monosemic verbs, while, in the second column, they refer to polysemous verbs only (60% of the total, with an average of 3 senses per verb). Let N be the number of target WCPs, A the number of topmost attempted answers (whether correct or wrong) and C the number of topmost correct answers: then recall is $\frac{A}{N}$, and precision $\frac{C}{A}$.

5.2 Experiment two

EXP2 covers a subset of the Italian results in the ROMANSEVAL competition, namely those obtained on the basis of lexico-semantic information only (as opposed to subcategorization information: for an overall discussion of ROMANSEVAL results see Federici, Montemagni and Pirrelli 1999). The corpus of target contexts consists of 665 verb occurrences of 20 polysemous verbs. The selected verbs show different degrees of polysemy: the number of senses ranges from the 16 senses of *passare* 'pass' to the two senses of *prevedere* 'foresee'; on average, each verb has five different senses. Unlike EXP1, target contexts include patterns already attested in EB or containing words unattested in EB.

For this experiment, SENSE EB was partitioned into two sets. The first partition is the same for all 20 target verbs and contains functionally annotated WCPs with no indication of sense (neither on the verb nor on the noun). It includes 17,359 verb–noun patterns (6201 with subject and 11,148 with object) automatically extracted from all verb entries of both the bilingual Collins dictionary and a monolingual Italian dictionary (Garzanti 1984). This set of patterns represents the typical usage of 3858 different verbs. Verbs are exemplified through a sparse number of WCPs, on average 4.5 per verb. We will refer to this partition as the 'unsupervised task'. The second partition, consisting of patterns illustrating the sense distinctions of the target verb, is much smaller and varies for each target verb W. It contains, given W, a comparatively small number of WCPs (28.3 on average) exemplifying an average of 5.4 contexts for each of W senses. This small subset of sense-tagged contexts can be viewed as a sample of sense seeds, according to Yarowsky's definition (Yarowsky 1995).

Given a target context T and the ambiguous verb W, the task facing SENSE in this experiment is to settle on a sense-tagged context in the sample of sense seeds of W such that it functions as the transfer term in a proportion where the two remaining terms (including the pivot) are looked for in the unsupervised tank. This requires the following reformulation of Step 2 in section 4.3:

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

Table 6. *An analogical proportion in EXP2*

'light–candle'	:	'switch_on–lamp'	=	'illuminate–candle'	:	'illuminate–lamp'

t_1	t_2	t_3	t_4
PRED: *accendere*	PRED: *accendere*	PRED: *illuminare*	PRED: *illuminare*
S_PRED: undef	S_PRED: undef	S_PRED: 1	S_PRED: **?**
SYNT_FUNCT: obj	SYNT_FUNCT: obj	SYNT_FUNCT: subj	SYNT_FUNCT: subj
ARG: *candela*	ARG: *lampada*	ARG: *candela*	ARG: *lampada*
S_ARG: undef	S_ARG: undef	S_ARG: undef	S_ARG: undef

Table 7. *Results of EXP2*

	Test corpus	Test corpus **without** known WCPs and unknown words
RECALL	73.7%	77.9%
PRECISION	79.0%	67.6%

Step 2-bis If there are (a) one term t_3 in the sample of sense seeds, and (b) two terms t_1 and t_2 in the unsupervised tank such that they are proportionally-related to T as t_4, then the sense of W in T is assigned according to condition 1′) in section 4.2; all possible solutions are eventually ranked.

Step 2-bis uses underspecified proportions such as the one in Table 6, intuitively less constrained than proportions 1 and 2 above. The semantic relatedness of nouns in the proportion is no longer proved to be S_PRED preserving, but more simply PRED preserving. Results are given in Table 7, where, once more, figures of recall and precision refer to the topmost ranked sense only. In the first column, figures refer to the results obtained for the whole test corpus of 665 contexts. Figures in the second column report the results for 412 target contexts only, obtained by discarding WCPs already known or containing unknown words.

It is useful to compare EXP2 results in the first column of Table 7 with the baseline obtained by assigning, for each test verb, the most frequent of its senses to all its WCPs in the entire test corpus. The baseline yields 100% recall and 59% precision. Precision of SENSE is significantly higher than in the baseline, at the price of lower recall. The scatter plot of Figure 1 compares precision figures for SENSE

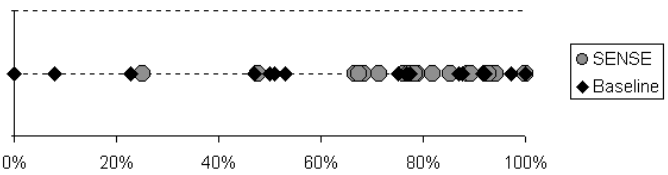


Fig. 1. Scatter plot of per-verb precision in SENSE and Baseline.

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

and the baseline more analytically, for each of the 20 test verbs. Unlike the baseline, whose precision ranges between 0% and 100%, the vast majority of test verbs (18 out of 20) are correctly disambiguated by SENSE with a precision higher than 67%.

5.3 Comparing experiments 1 and 2

EXP1 and EXP2 yield considerably different results, due to the different EBs, the diversity of the test protocols and of the purposes they were meant to serve, namely: evaluation of the plausibility of assumption 3 in section 3 through maximally-constrained proportions on the one hand (EXP1), and evaluation of the effectiveness of the same assumption through minimally-constrained proportions on the other (EXP2). A comparison of recall and precision in the two settings is meaningful only in relation to the second column of Tables 5 and 7.

What EXP1 shows is that assumption **(3)** is notably accurate (80.4% precision), albeit at a cost in terms of recall, with 33.7% of input contexts which cannot be related to known examples through viable proportions. The cost is not extravagant, given the comparative small size of EB in the first experiment. Precision in the face of sparse overt evidence shows SENSE capacity of using series of data, as opposed to examples in isolation. This makes it possible to capitalise, on a principled basis, on (i) correlation of selectional restrictions with classes of predicates and (ii) systematic regularities in meaning shifts, thus making the most of sparse training data.

In EXP2, we tried to improve on recall by increasing the size of EB and by weakening assumption **(3)**. This leads to the use of less constrained proportions, with non disambiguated pivots. An important consequence of this move is that sense assignment is based on a considerable amount of unsupervised data, through a sample of sense seeds. Recall is significantly higher than in EXP1, yet at the cost of being less accurate. It should be emphasised that the inferential algorithm of EXP2 is strongly unsupervised, as opposed to weaker versions of unsupervised WSD such as the one proposed by Yarowsky (1995). There, initial samples of sense seeds, bigger than those used for EXP2, are incrementally updated through accretion of further disambiguated contexts. Here, the sample of sense seeds is never incremented, and proportional trasferring is carried out entirely on the basis of untagged contexts, thus lending the most direct support to the idea of proportionally-based word sense similarity.

A comparison of EXP1 and EXP2 with results obtained through other approaches to WSD, although very interesting in theory, is not possible in practice, due to substantial differences in the test conditions, e.g. typology and granularity of the senses of target words, types of text, etc. To the best of our knowledge, the forthcoming issue of *Computers and the Humanities* on SENSEVAL/ROMANSEVAL, where EXP2 is described, is the first attempt to provide a common basis for comparative evaluation of WSD systems.

6 Concluding remarks

WSD is a linguistically multidimensional task and calls for powerful means of establishing multidimensional relations among attested usages of word senses. We

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

show that multidimensional analogy-based proportions have this potential. They prove to be able to ground inter-contextual similarity on sense invariance, to the effect that contexts are judged similar only in relation to their capacity of preserving the sense of a target word. This is certainly more conservative than grounding sense invariance on context similarity taken at face value (as done by most WSD systems we know of) and is bound to lower recall. SENSE is able to make up for its conservative bias by exploiting systematic proportional relations among patterns in EB. We also intend to provide SENSE with a long range memory for the algorithm to capitalise on previous sense assignments and take advantage of the 'one sense per discourse' principle.

Multidimensional proportions offer the further advantage that one can possibly weaken some of the constraints they enforce with no dramatic degradation in performance. This is the key to an understanding of the relative success of EXP2. Sense preservation is no longer checked, but still contexts are related in a meaningful way, so that a handful of sense-tagged contexts (sense seeds) can be extended reliably through a set of untagged data (unsupervised tank). The practical consequences of this are far reaching. For a given W, one can 'plug' into EB different typologies of sense subdivision, with different degrees of granularity, to disambiguate a corpus with respect to the required sense distinctions. Moreover, also the unsupervised task can either be extended through new WCPs extracted from unrestricted texts, or specialised through addition of domain-specific contexts.

It should be appreciated that the algorithm described in these pages is both language-independent and applicable to any part of speech, provided that relevant WCPs are available. We also expect it to be usable as a discovery procedure for sense identification and automatic induction of semantic classes. Preliminary experiments have so far borne out this expectation.

References

Abe, N. and Li, H. (1996) Learning word association norms using tree cut pair models. *Proceedings of the XIII International Conference on Machine Learning*.

Collins (1985) *English–Italian Italian–English Dictionary*. Collins Giunti Marzocco, London Firenze.

Federici, S. and Pirrelli, V. (1994) Linguistic analogy as a computable process. *Proceedings of NeMLaP*, Manchester, UK, pp. 8–14.

Federici, S., Montemagni, S. and Pirrelli, V. (1996) Analogy and relevance: Paradigmatic networks as filtering devices. *Proceedings of NeMLaP*, Ankara, Turkey, pp. 13–24.

Federici, S., Montemagni, S. and Pirrelli, V. (1997) Inferring semantic similarity from distributional evidence: an Analogy-based approach to Word Sense Disambiguation. *Proceedings ACL/EACL Workshop 'Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications'*, Madrid, Spain.

Federici, S., Montemagni, S. and Pirrelli, V. (1999) ROMANSEVAL: results for Italian by SENSE. In: M. Palmer and A. Kilgarriff (eds.), *Computers and the Humanities*, Special issue on SENSEVAL.

Garzanti (1984) *Il Nuovo Dizionario Italiano Garzanti*. Garzanti, Milano.

Ide, N. and Véronis, J. (1998) Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art. *Computational Linguistics* **24**(1): 1–40.

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press

Kelly, E. F. and Stone, P. J. (1975) *Computer Recognition of English Word Senses*. North–Holland, Amsterdam.

Lin, D. (1997) Using syntactic dependency as local context to resolve word sense ambiguity. *Proceedings of ACL/EACL–97"*, Madrid, Spain, pp. 64–71.

Melamed, D. (1997) Measuring semantic entropy. *Proceedings of SIGLEX Workshop on Tagging Text with Lexical Semantics: why, what and how? ANLP'97*, Washington, DC.

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

Montemagni, S. and Pirrelli, V. (1998) Augmenting WordNet-like lexical resources with distributional evidence. An application-oriented perspective. *Proceedings COLING–ACL '98 Workshop on 'Usage of WordNet in Natural Language Processing Systems'*, Montreal, Canada.

Pirrelli, V. and Yvon, F. (1999) The hidden dimension: paradigmatic approaches to data-driven NLP. *J. Experimental & Theoretical AI* (to appear).

Vossen, P., Ide, N. and Greenstein, D. (eds.) (1998) *Computers and the Humanities. Double Special Issue on EuroWordNet* **32**(2–3).

Yarowsky, D. (1995) Unsupervised word sense disambiguation rivaling supervised methods. *Proceedings ACL '95*, Cambridge, MA, pp. 189–196.

https://doi.org/10.1017/S135132499900217X Published online by Cambridge University Press