UNIVERSITY OF MURCIA

International Journal of English Studies

IJES

www.um.es/ijes

The Treatment of Word Sense Inventories in the 'LACELL WSD Project'

MOISÉS ALMELA[1]
LACELL RESEARCH GROUP – UNIVERSIDAD DE MURCIA

ABSTRACT
The WSD community has long debated whether the criteria for representing polysemy in general purpose dictionaries meet the specific demands of sense disambiguation tasks. Concern is growing that pre-defined sense inventories might not adjust well to the needs of WSD, because word occurrences can rarely be paired with rigid sense classes in a one-to-one fashion. A second cause for concern is the level of sense granularity adopted in conventional dictionary entries. Fine-grained distinctions can be useful for a dictionary user but complicate the design and evaluation of WSD systems in a way that is often unnecessary. As a result of these objections, many experts have voiced the opinion that dictionaries are not adequate sources of sense inventories for WSD. However, the problem of word sense overlaps can also be resolved by modifying the way in which dictionary entries are processed by WSD programs. This is the solution applied in the LACELL WSD system. The algorithm selects simultaneously two or more dictionary senses if the context does not allow sufficient discrimination between/among them. This article explains the underpinnings of such proposal, as well as discussing some advantages and disadvantages.

KEYWORDS: WSD, Lexicology, Lexicography, Corpus Linguistics, Computational Linguistics.

RESUMEN
En el ámbito de la investigación sobre desambiguación léxica automática (WSD), se ha venido debatiendo largo tiempo acerca de la adecuación de los modelos de análisis polisémico empleados en los diccionarios de carácter general. En particular, existe una creciente preocupación en torno a los problemas generados por la utilización de inventarios de acepciones léxicas, ya que no es frecuente hallar correspondencias biunívocas entre los usos de una palabra en contextos específicos y las clases semánticas preestablecidas en la entrada léxica. Además, se duda de que el nivel de granularidad semántica aplicado en la lexicografía convencional sea el más adecuado para las necesidades específicas de la WSD. Como consecuencia de estas objeciones, algunos expertos han llegado incluso a sugerir que los sistemas de WSD deberían prescindir de las entradas de los diccionarios como sus principales recursos léxicos. Sin embargo, hay otras soluciones alternativas, entre ellas

[1] ***Address for correspondence***: Moisés Almela Sánchez, Dept. de Fil. Inglesa, Univ. de Murcia, 30100 Murcia; Tel.: +34 868 883212; e-mail: moisesal@um.es.

modificar el modo como los programas de desambiguación automática procesan los repertorios de acepciones léxicas listados en las entradas de los diccionarios. El sistema de WSD creado por el Grupo LACELL está diseñado para seleccionar simultáneamente un grupo de subacepciones en aquellos casos en los que el contexto no proporciona suficiente información para discriminar entre ellas. En este artículo se explican los fundamentos de dicha propuesta, además de comentar algunas de sus principales ventajas e inconvenientes.

PALABRAS CLAVE: WSD, Lexicología. Lexicografía, Lingüística de corpus, Lingüística computacional.

I. WORD MEANING: AN INTERDISCIPLINARY CONCERN

In 2004, the LACELL Research Group at the University of Murcia embarked on the project of developing an algorithm for automatically disambiguating word senses, thus becoming one of the first Humanities Departments in Spain to undertake research in this field. Hitherto, WSD research had been almost the exclusive remit of information technologists. Input from the realm of linguistics has gradually become more and more necessary, as WSD experts have come to realize that further development is being hindered by the lack of adequate models of lexical meaning. In an era of rapidly advancing engineering technologies, when the processing capacity of computers falls little short of ideal, the most serious obstacles to automatic disambiguation of senses are posed not by a want of technical achievements, but rather but by a lack of resolution over perennial lexicographic issues such as the fuzziness of word sense boundaries or the relationship between meaning and usage. The project undertaken by the LACELL Group emerged as an attempt to base WSD developments on innovations in the area of lexical semantics.

WSD and lexical semantics share a preoccupation with the phenomenon of polysemy, that is, the modification of word meaning in context. Unfortunately, lexical semantics has remained reluctant to put its models in tune with the demands from WSD, and conversely, WSD experts have long underrated the possibility that theoretical-linguistic innovations might result in performance increases of sense disambiguation systems. It has not been till recently that leading scholars in the WSD community have taken a more optimistic stance on the impact that new theoretical models of word meaning might have on WSD performance. In their introduction to a recent monograph on WSD, Agirre and Edmonds suggested that one of the ways to achieve fundamentally new research results in the future is to "become more theoretical [...], to work on WSD embracing more realistic models of word sense (including non-discreteness, vagueness, and analogy), thus drawing on and feeding theories of word meaning and context from (computational) lexical semantics and lexicography" (Agirre & Edmonds, 2006: 17). However, this requires cooperation on the part of lexicologists and semanticists, and, as the same authors regret to inform, "WSD has never really found a home in lexical semantics", despite the fact that "word meaning is at the heart of the problem" (Agirre & Edmonds, 2006: 2).

Considering that one of the objectives of lexical semantics is to "understand the relationships between 'word', 'meaning', and 'context'" (Agirre & Edmonds, 2006: 2), it is indeed surprising, as well as undesirable, to find that lexicological contributions to WSD systems have been an exception rather than the rule. Traditionally, WSD has been allied more closely with lexicography than with theoretical lexical semantics. Agirre and Edmonds point out two plausible reasons for this. The first one has to do with the adequacy of the descriptive mechanisms employed in each of these fields. Lexical semantics, according to Agirre and Edmonds (2006: 2), "has always been more concerned with representational issues and models of word meaning and polysemy so far too complex for WSD". More suitable for the needs of WSD is the "intuitive premise that word uses group into coherent semantic units" — an intuition that is not as widely embraced by semanticists as by modern lexicographers (Agirre & Edmonds, 2006: 2). The second reason mentioned is the affinity of WSD methodologies with the empirical corpus-based approaches that have become popular in modern lexicography.

In sum, the message that WSD experts have sent to lexicologists can be briefly summarized as follows. In principle, research conducted in the area of lexical semantics can be integrated within the interests of WSD as long as it meets two conditions: firstly, it has to be modelled on the idea that usages of a word will tend to cluster together in groups of semantic proximity; and secondly, its statements must be underpinned by statistically-aided observations of how words and their senses behave in representative samples of discourse. The rationale for the first recommendation is to get around the ongoing problem of word sense boundaries, an issue that will be discussed below in section II. As for the second recommendation, it must be recognized that corpus-based lexicology and lexicography are now well on the way towards providing empirically sound models of the syntagmatic behaviour of words, and this behaviour, in turn, can be mapped onto semantic descriptions, as Firth predicted in his formulation of the postulate of meaning by collocation.

II. THE PROBLEM OF WORD SENSE DISCRIMINATION

"The trouble with word sense disambiguation is word senses", reads the opening line of an article written by Kilgarriff (2006: 29). Indeed, a major pitfall of sense disambiguation has been the lack of principled criteria for drawing the boundaries between different senses of a word: "there are no decisive ways of identifying where one sense ends and the next begins" (Kilgarriff 2006: 29). In turn, the lack of criteria for delimiting different word senses can be seen as a symptom that the category 'word sense' is ill-defined. The problem, of course, is not new, but as more and more disciplines and lines of research are becoming interested in arriving at a consensual definition of what a word sense actually is, the lack of such consensus becomes more and more obvious, and the consequences less bearable. As two WSD experts remarked, "the problem of sense division has been an object of discussion

since antiquity [...], but the lack of resolution over 2,000 years is striking" (Ide & Véronis, 1998: 22).

Recently, the problem of word sense division has surfaced in discrepancies among human annotators, particularly between so-called 'lumpers' and 'splitters' (Fellbaum et al. 2005). From an experimental study Véronis (2003: 287) concludes that "interannotator disagreement is very low in a straightforward sense tagging task using a traditional dictionary. For some words, agreement was no better than chance". The evidence of annotation disagreements has precipitated a wave of dissatisfaction with the techniques of the lexicographic tradition on which natural language processing (NLP), and WSD tasks in particular, had relied for so many years. Ide and Wilks (2006) have even gone so far as to hint that lexicographic expertise in WSD is altogether redundant. There is a general feeling that the pressing demands of WSD applications have finally put conventional lexicography to the test, after several centuries of unchallenged authority, and discontent with the traditional model of a semantic unit is growing. The conventional concept of the word sense as a discrete entity has come under attack, and the representational models of word polysemy as a fixed or static inventory of senses is now called into question.

Reactions to the problem of word sense have moved in three directions, which are briefly described below:

(1) *Relativization*. Some authors have adopted a skeptical, nominalist stance on the concept of word sense. This concept, explains Kilgarriff (1997: 108), is not "sufficiently well-defined to constitute a workable basic unit of meaning". On his view, word senses are not be included as objects within the ontology of linguistics, because the limits between different senses are neither real or unreal, but only convenient or inconvenient, that is, useful for a particular task or not: "word senses are only ever defined relative to a set of interests" (Ibid.).

(2) *Dynamization*. The standard lexicographic treatment of word senses has presented them as lists of discrete entities. Likewise, in WSD, sense labels are in most cases drawn from pre-defined sense inventories—most typically those found in lexical resources such as WordNet or machine-readable dictionaries. However, since the mid nineties some scholars have argued that a lexicon endowed with the capacity for 'generating' senses outperforms a lexicon in which senses are merely 'enumerated' (Pustejovsky, 1995). The Generative Lexicon (GL) theory is equipped with highly sophisticated rule sets for composing actual senses based on the interaction between word meaning and context. Fellbaum et al. (2005) explain that, by replacing pre-defined sense lists with mechanisms of semantic composition, such as those of the GL, it might be possible to solve one of the main problems with which WSD systems are faced, to wit: the lack of one-to-one correspondences between dictionary senses and actual uses of a word in a corpus. Often, the way in which a word is used in the discourse overlaps boundaries between senses that are separated in the dictionary.

(3) *Hierarchization*. The grouping of subsenses into main senses has been a well-known strategy in lexicography, but its applicability in WSD methods has not been thoroughly explored. Basing on findings from a translingual empirical study, Resnik and Yarowsky (1999) put forward a four-pronged sense hierarchy, whereby the lowest degree of granularity is represented by the 'homograph', and the highest degree of granularity is represented by the 'subsubsense' level. In between these poles lie two intermediate categories, namely, the 'major sense' level and the 'subsense' level. However, it is essential to point out that discrimination among subsubsenses or even among subsenses is not possible in all types of contexts. Not all word occurrences can be assigned the same level of sense granularity. The failure to ascribe the correct sense to a word is often the result of an attempt to disambiguate at a higher level of granularity than required. Therefore, a sound hierarchization of senses in the lexical entry is only of limited use if the WSD algorithm is unable to decide on the appropriate level of granularity for each word occurrence. One of the main objectives of the LACELL project on WSD has been to devise a system capable of identifying cases where the context drops any distinction among subsubsenses, as well as cases where various subsenses are activated at the same time. The idea is that the performance of WSD systems can be increased if the algorithm has the capacity to stop the disambiguation process at the maximum level of granularity allowed by the context.

The first proposal, namely, to abandon any attempt at finding objective (task-independent) criteria for sense delimitation, is less justified for coarse-grained than for fine-grained distinctions. Arguably, inter-annotator disagreements do not affect all levels of sense discrimination alike. It is reasonable to hypothesize that the degree of difficulty and discrepancy is directly related to the granularity of an inventory. In fact, "the most often-cited obstacle to correct assignment of pre-defined senses concerns granularity" (Ide & Wilks, 2006: 52). Discrimination among underspecified senses raises less controversy and intersubjective divergencies than discrimination among highly specific subsenses. It has been suggested in the literature that finer-grained distinctions should be "collapsed to their highest parent" (Ide & Wilks, 2006: 53) as a means to avoid the problems posed by mismatches between actual uses and pre-defined inventories. Disagreements as to which sense applies to each use tend to be sharper as the granularity of the sense classes increases. Ide and Wilks (Ibid.) argue that "homograph-like" distinctions can be drawn neatly, even without recourse to lexicographic expertise, and that most WSD tasks do not require sense distinctions beyond the homograph-like level. Besides, they report that "all successful WSD has operated at what [...] we could call the homograph rather than the sense level" (Ibid.).

The success of automatic disambiguation at low-granularity levels can be advanced as an argument for the objective reality of broad sense classes. It would be hardly possible to achieve high disambiguation rates with such senses if they were not characterized by distinct

sets of (co-textual) properties. This is consistent with some findings from psycholinguistics. Klein and Murphy (2001, 2002) have found behavioural evidence that polysemous senses (or rather: certain types of polysemous senses) are stored separately in the mental lexicon, and that their processing resembles that of homonyms. That is to say: senses are functionally and representationally distinct. Nevertheless, it is important to note that the aforementioned evidence is far from conclusive, as there are studies both in psycholinguistics and in cognitive neuroscience that point in a different direction. Frisson and Pickering (1999) find behavioural evidence suggesting that the different senses of a word are derived from a single representation, i.e. an underspecified core meaning. The hypothesis of separate lexical representation is also contradicted by neurolinguistic experiments conducted by Pylkkänen et al. (2006), but these authors suggest that theoretical interpretations other than the core meaning hypothesis are plausible too.

Thus, the question of how best to represent polysemous categories remains an unresolved debate across disciplines such as lexicography, psycholinguistics and NLP, inter alia. More specifically, the dilemma between storage and generation constitutes one of the most controversial issues under discussion, as well as one of the more promising research questions. As Klein and Murphy (2002: 568) remark, "what is most needed now is a better understanding of which aspects of word meaning are stored and which are constructed on the basis of context and pragmatic inference". Mutatis mutandis, the same debate is paralleled in lexicology and lexicography by the dialectics of 'enumerative' and 'generative' lexicon. GL adherents propose to minimize storage (representational) costs by maximizing the scope of rules of semantic composition. This linguistic programme dovetails with what –in the field of psycholinguistics– Frisson and Pickering (1999) called the 'underspecified account' of word sense mental processing. These authors describe the process as a top-down mechanism of successive refinement. The initial unit is an abstract semantic value (an underspecified meaning) with no commitment to any of the specific senses. It is only during the interaction with the context that the more specific features of meaning are incorporated into the word.

However, there is no such precept that demands a high level of semantic specification in each word occurrence. Vagueness is an option, and it results from situations in which no contextual specifications are homed in on the initial (underspecified) meaning: "It may not be obligatory to determine the exact sense if the context is vague with respect to the appropriate interpretation. As long as some semantic value can be assigned to an expression, without making the sentence semantically incongruous at that point (and as long as it is not a homonym), readers may not have to choose between the different senses of that word" (Frisson & Pickering, 1999: 1379).

The hypothesis of an underspecified core meaning can thus be credited with psycholinguistic plausibility, and this, in turn, can be interpreted as an argument for the theory of the GL. However, the application of this theory in the field of WSD must be

confronted with a number of challenges. Firstly, implementation of a WSD system based on the GL theory involves enormous computational costs, due to the complexity of the rules of semantic composition. Secondly, doubts have been raised concerning the scope of sense variations which the GL theory can predict. Kilgarriff (2001) denies that the GL can account for non-standard uses of words. Besides, the GL was originally conceived to explain 'logical polysemy', not to analyse 'contrastive ambiguity' (Pustejovsky, 1995: 29). This leaves outside the scope of this theory a host of sense distinctions at low granularity levels, such as *buck* = 'animal' / *buck* = 'money'. The GL is especially suitable for predicting fine-grained metonymic variations, such as *door* = 'physical object' / *door* = 'aperture'. However, WSD tasks require the resolution of variegated forms of sense extensions, not only of metonymies.

In view of the complexities and the limitations of the GL, the LACELL research group opted for a strategy of sense hierarchization in an enumerative lexicon. The system pursued two priorities: first, minimize computational costs, and second, deal with different degrees of sense granularity. Computationally, this model had the advantage of involving relatively low-dimensional feature space, because it was designed to run on raw co-textual data, hence dispensing with all kinds of enriched semantic representations. Disambiguating clues were based on the parameters of relative position and collocation alone. As regards the statistical technique, the use of discriminant function analysis made it possible to compute distances between each occurrence and each semantic class; for each meaning, the system determined the location of the point (group centroids) that represents the means for all variables (collocational data); then, for each case, the algorithm computed the distances (of the respective case) from each of the group centroids. Finally, cases were classified as belonging to the group (meaning) to which each of them was closest—for a more detailed description of how the algorithm works, cf. Sánchez et al. (2007).

One of the strengths of the system devised by the LACELL group was the possibility to automatically detect cases of 'subsense merger', that is, cases in which specifications of a major sense were not required, because the different subsenses could be activated simultaneously. The strategy employed was to compare the distance of each case to the different group centroids. In this way, it was possible to determine whether one case (word occurrence) had a similar or even identical distance to more than one group centroid. If the distance between a case and one of the centroids is significantly shorter than with all the other centroids, the word can be safely disambiguated into the sense represented by the centroid to which it is closest. However, if the case under analysis stands at a similar distance to two or more group centroids, the data can be interpreted as reflecting a lack of discriminability between subsenses. This problem, and how the LACELL WSD system copes with it, shall be the topic of the next section.

III. VAGUENESS VS. AMBIGUITY

Above, it has been argued that speakers do not always have to choose between different readings of an underspecified meaning, and that sometimes the correct interpretation of a word occurrence is a vague semantic value (Frisson and Pickering 1999). This reminds us that the notions of 'ambiguity' and 'vagueness' have often been used in confusing ways in the literature. Ambiguity is a property of (polysemous) lexical units, while vagueness is a property of contexts. In the language system, when considered in abstraction from the particular contexts of use, lexical units are ambiguous, because they have a potential for denoting more than one semantic category. Disambiguation takes place in actual contexts of use, as the interaction of the word with other elements indicates which features of the meaning potential are congruent with other categories and relations expressed in the environment. However, not all contexts of use have the same power of sense discrimination. In high-discriminating contexts, there is only a specific subsense or even a subsubsense of the word to cohere with the environment; in low-discriminating contexts, by contrast, several subsenses can be selected at the same time without negative effects on text cohesion. In such cases, discrimination beyond the major sense level is irrelevant.

It is therefore necessary to distinguish between cases of ambiguity and cases of vagueness. Ambiguity is a situation in which two or more senses compete for selection. This competition often takes place at low-granularity levels. For instance, it is a typical strategy in puns to activate simultaneously senses which belong to different conceptual domains, as in the following pro-animal rights slogan: *One fur hat, two spoilt bitches*. Here, the potential ambiguity of the words in the second noun phrase is not resolved. Of course, this is a conscious strategy intended to catch the reader's attention. The aforementioned example exploits the resources of contrastive ambiguity for rhetorical purposes. In cases of contrastive ambiguity, like the one just mentioned, the semantic domain of one item cannot be selected independently of the domain of one or more items in the environment (Pustejovsky, 1995).

The phenomenon of vagueness is essentially different from ambiguity. Vagueness can be defined as a situation in which the correct interpretation of a word is located at a low-granularity level and no specification of subsenses –let alone of subsubsenses– is required. In a vague context, the differences among exact senses are neutralized, and no further specification is homed in on the core meaning. A vague meaning is thus the result of reducing or even cancelling fine-grained sense distinctions. Vagueness plays an important function in communication, because speakers often prefer to avoid commitment with any of the exact interpretations of the word.

As was mentioned in the previous section, discriminant function analysis is the statistical technique which the LACELL WSD system uses to capture degrees of proximity between word occurrences and sense classes. A graphical display of results obtained with this type of analysis is shown in figure 1. Here, each of the circles represents a case (word occurrence),

while the squares represent group centroids (sense classes). As can be seen in the figure, some word occurrences are very close to one of the centroids and distant from all the others. In some other cases, however, the actual use of the word stands at a similar distance to more than one centroid. This reflects the fact that more than one specific sense can be selected at the same time and is a plausible interpretation of the word.

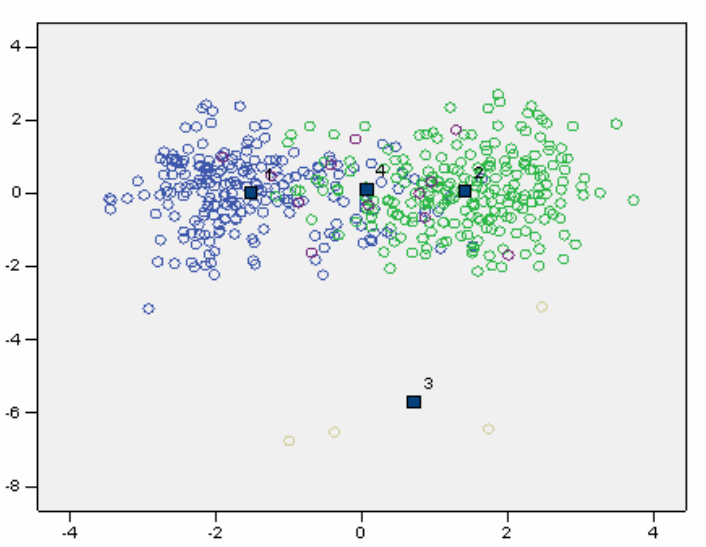

Figure 1. Discriminant function analysis of word senses.

Let us now explain how this works with a particular example. Figure 2 shows the results of the disambiguation of a sentence with *heart*, as displayed in the interface used by the LACELL WSD system. The sentence was typed on the top side of the screen. The bottom side of the screen shows the output, which in this case yields a set of five senses (the definitions are based on the Merriam Webster Dictionary, 2000). Distinctions among these senses are rather fine-grained, which explains why the program does not discard any of them. The use of the noun *heart* in this sentence (*Her work is a passion which she holds close to her heart*) activates meaning components that are scattered over the senses selected: 'personality', 'attributes', 'thought', 'consciousness', 'feeling', 'disposition', 'spirit'. There is no one-to-one correspondence between this use of the word *heart* and one of the senses listed in the dictionary entry. However, the group of five senses selected is sufficiently homogeneous to provide a coherent representation of the way in which *heart* is to be interpreted in this context. All five senses make allusion to a person's psychological constitution. The fact that *heart* is used here in a phraseological context (*close to one's heart*) does not prevent it from being assigned a sense and a referential function, because the noun retains a certain degree of semantic autonomy (a similar example will be analysed in the next section). In sum, *heart* has here a relatively vague semantic value, because its

correct interpretation requires the neutralization of a series of fine-grained distinctions, such as that between 'personality' and 'consciousness', or between 'attributes' and 'disposition', among others.

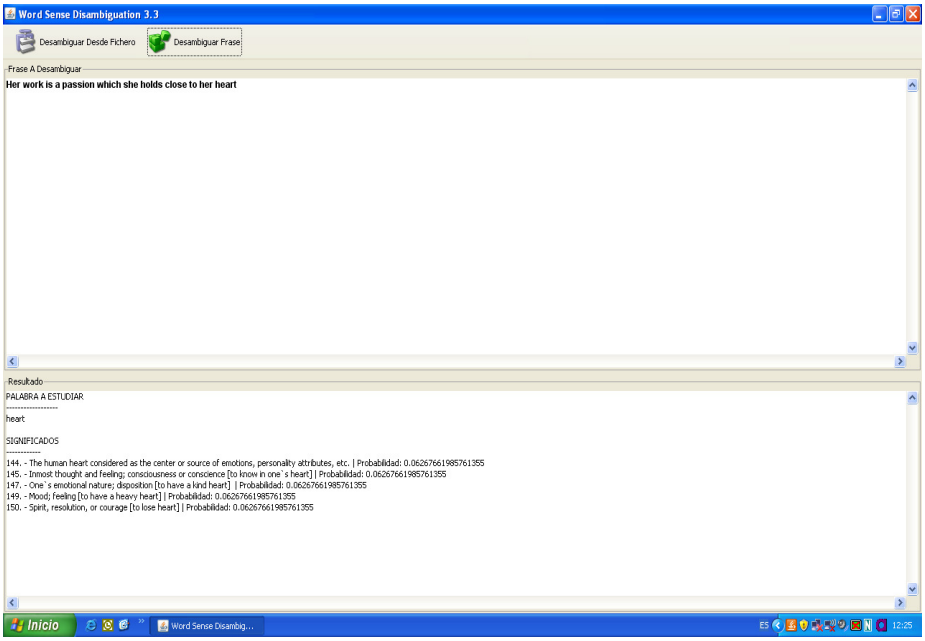

Figure 2. Representation of a case of 'vague semantic value' in the LACELL WSD interface.

Crucially, the aforementioned example is a case of vague interpretation, but not of ambiguity. It is essential to underline the fact that the potential ambiguity of *heart* is resolved at its basic (low-granularity) level. Thus, the program does not select in this context other senses of *heart*, such as 'hollow, muscular organ', or 'something like a heart in shape', or 'a red figure of a heart stamped on a playing card', or 'the human heart considered as the central part or core of something'. Ambiguities of the type that are usually exploited in puns, i.e. those which involve competition between senses from different domains, are resolved by the algorithm in the case indicated above. In the sentence *Her work is a passion which she holds close to her heart*, the word *heart* does not receive incorrect interpretations such as those referring to the body part or to the geographical centre of a place. The program has proven able to detect that the use of *heart* in this context refers to some phenomenon of the psychological sphere, rather than having a bodily or geographic sense. However, finer-grained distinctions are irrelevant, because the context activates various psychological senses of *heart*.

IV. SENSE NETWORKS

Sometimes the result of automatic disambiguation highlights the mechanisms of sense extension, rather than the mere grouping of subsenses under a common superordinate category or notion. Consider, for example, the sentence analysed in Figure 3, with the Spanish noun *mano* (= hand) as the unit under scrutiny: *A partir de ese momento rigió la justicia de la mano propia, era matar o morir.* The program has selected five senses for this occurrence of *mano*, which I shall briefly refer to as (1) human body part, (2) meat, (3) power, (4) agency/power, and (5) skill, respectively (the expanded definitions of the senses provided in the WSD interface –see Figure 3– are quoted from the *Gran diccionario de uso del español actual*, SGEL (2001)).

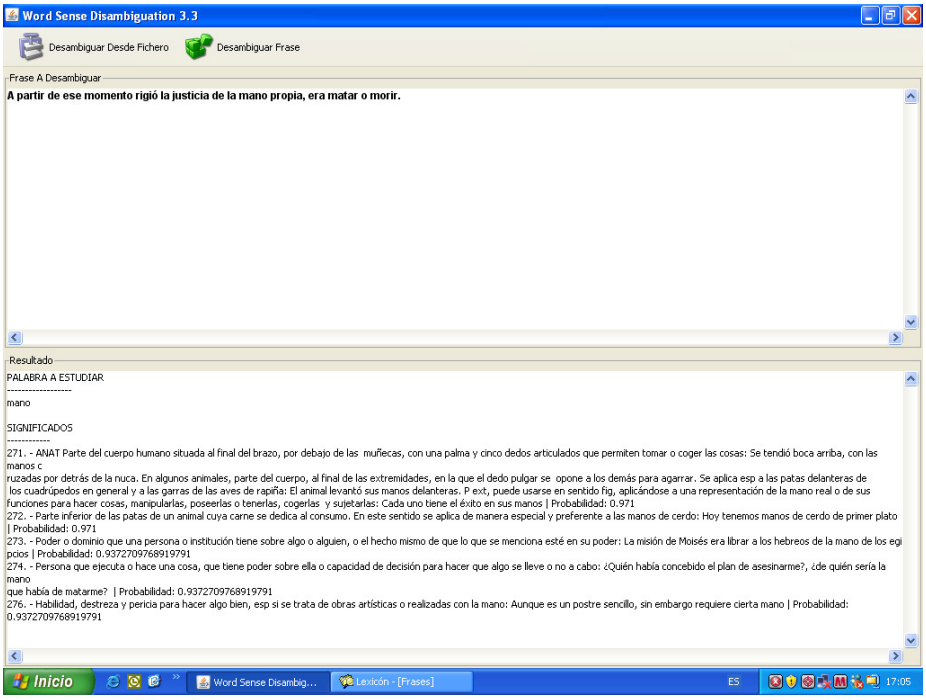

Figure 3. Interference of a sense network in the disambiguation of *mano*.

In principle, the sense that best captures the use of *mano* in this sentence is the one that includes the features 'agency/power', that is number (4) above. However, the algorithm has erroneously selected number (1) and (2) as the most probable ones. This time the program has failed to accurately disambiguate the word under analysis. Strictly speaking, this is not a case of subsense merger (as in Figure 2), because the five possible interpretations selected do not even form a 'taxonomic' category. The senses 'body part', 'power', 'agency' and 'skill' cannot be grouped together according to class membership. However, they can be said to join in the same 'thematic' and 'ad hoc' categories, i.e. categories whose members

are linked together by co-occurrence, association, functional relations, or goals—for a discussion of different ways of classifying things, according to categorization research, cf. Klein and Murphy (2002). Senses (3), (4) and (5) extend pragmatically from various aspects of sense (1), because the organ 'hand' is conceptualized in such a way that symbolizes the ability of individuals to wield power, carry out actions, or perform difficult tasks. Just as the organs 'head' and 'hand' are strongly associated with the 'intellect' and the 'passions', respectively, the organ 'hand' is strongly associated with the concept of 'action'.

Hence, it can be argued that in Figure 3 the program has identified a subset of motivated sense extensions, rather than disambiguating the word sense. The algorithm has not picked out the one sense that matches the use of *mano* in this sentence, but it has selected a network of senses that are pragmatically as well as phylogenetically related to the correct interpretation. It could be significant to note that senses extending from other aspects of the meaning of *mano* have not been selected. The program has not grouped together senses that are not directly related to the features 'control, agency, power', such as 'hands of a clock or watch', 'set of cards', or 'coat of paint', among others. This raises an interesting question, to wit: whether the underlying phylogenetic relationship between a primary sense and its derived senses, or between a source domain and a target domain, leave any trace in their collocational relations. In fact, previous corpus-based research has shown that some collocates of polysemous words can be associated with both literal and figurative senses of the node (Deignan 2005).

A case in point is the use of the verb *take* in combination with the 'body part' sense of *hand*, or as part of the phraseological expression *to take a hand (in something)*. In idioms that are motivated and analysable, nouns can be referential, because individual parts of the expression retain a certain degree of semantic autonomy (Stathi 2007). Hence, it seems appropriate to assign the senses 'control, power, authority' and 'agency, influence' to the occurrence of the word *hand* in the sentence *The authorities are taking a hand*. However, alongside these senses, the program has also selected the primary sense 'human body part' (see Figure 4), which is obviously incorrect. If we wonder why this has been so, we might find a plausible answer in Figure 5, which disambiguates the use of *hand* in the sentence *He nodded, and took her hand and kissed it*. Here, *take* is used as a verbal collocate of *hand*, and the program selects exactly the same set of three senses as in Figure 4, but the correct interpretation is only the primary sense, namely, human body part.

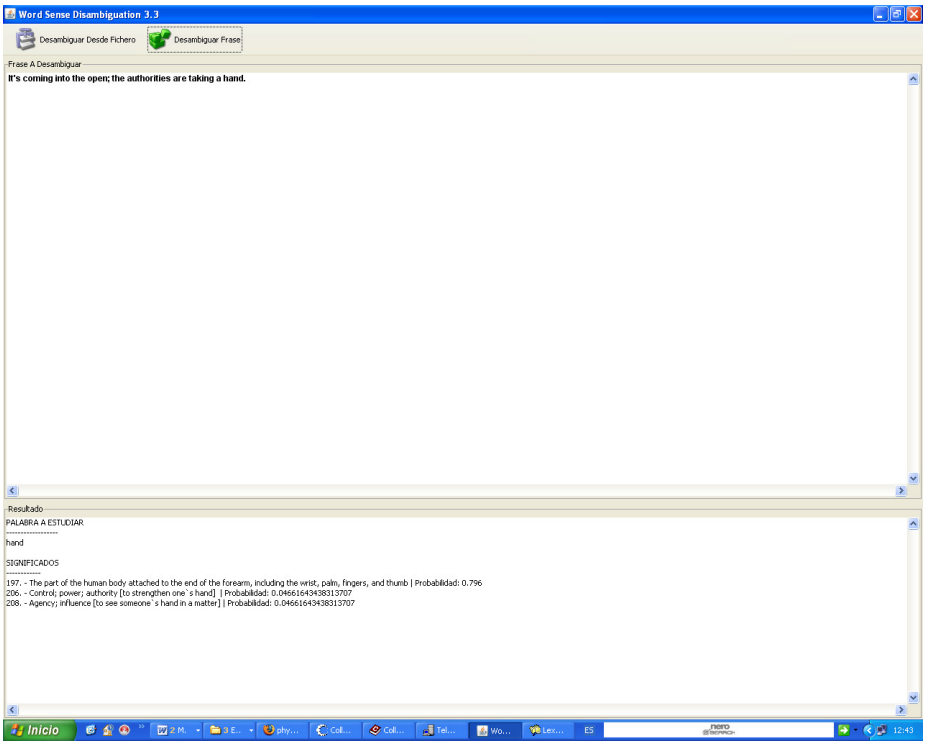

Figure 4. Interference of a sense network in the disambiguation of *hand*: failure to disambiguate the figurative sense.

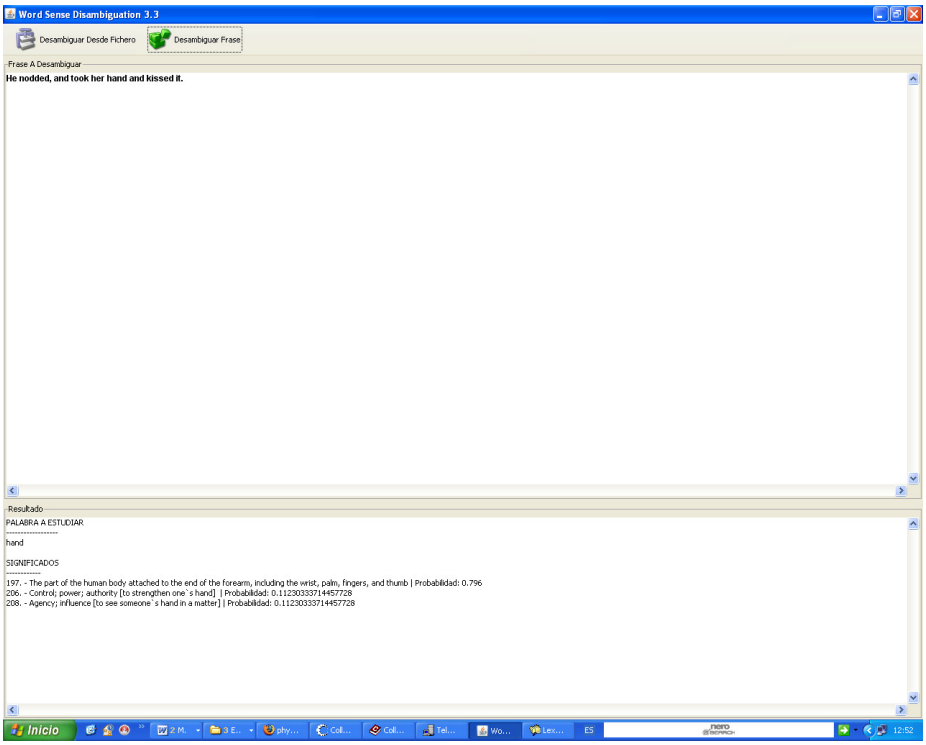

Figure 5. Interference of a sense network in the disambiguation of *hand*: failure to disambiguate the literal sense.

Apparently, collocation with *take* induces the program to automatically select a network of phylogenetically related senses, including a pair of figurative subsenses and the literal sense (or source domain) from which they are derived. The cause of this erroneous disambiguation, where the program selects a wider range of senses than required, lies in the 'versatility' of *take*. This verb is associated with more than one conceptual domain of the noun *hand*. This reveals a more general fact: some collocates of the figurative uses of a word are so to say 'inherited' from the syntagmatic contexts of the primary meaning.

The fact that some items collocate alternately with literal and figurative senses of the same word poses a serious problem to a WSD system such as the one devised by the LACELL group, which relies heavily on collocational information. The implication is that some collocates are not sufficient to disambiguate even at a level of coarse-grained sense distinctions, such as those established between literal-concrete and figurative-abstract interpretations of a word. There are two possible solutions to this problem. Firstly, the WSD system can be rendered more effective if it is enriched with deep semantic representations, for instance annotation of semantic classes. However, this also implies an increase in computational costs, as was discussed in section II. The rationale for relying mainly on surface collocational information in the LACELL WSD system is that, in doing so,

computational costs are kept to minimum. A second possible solution would be to enhance the parameters of contextual variation so as to include colligations (i.e. patterns involving association with specific word classes, or with specific grammatical items). Thus, in the case of *take* and *hand*, the ambiguity of the lexical collocation is usually resolved by the colligation. The verbal collocate is associated both with the literal and the figurative sense, but the determiners differ. The use of a possessive determiner between *take* and *hand* favours the selection of the body part sense, while the occurrence of the indefinite article in the same position favours the recognition of the idiomatic expression, and consequently, the activation of the figurative senses related with 'authority' and 'agency'.

However, leaving the worries of WSD tasks for a moment, the problem described above is of special interest to the semanticist, for it suggests that mechanisms of sense extension are realized lexically (collocationally) at the surface level of discourse—in other words, shared collocates among different senses of a word can be regarded as traces of the common origin of such senses. Of course, more empirical research would be required before this hypothesis can be substantiated. The LACELL WSD system can be used as a means of further testing this supposition. By observing and analysing which groups of senses are offered as the most probable interpretations of each token, it is possible to determine in which cases the senses selected reflect a network of pragmatically and/or phylogenetically connected categories (Figures 3-5), rather than clustering around a single taxonomic category (Figure 2).

V. CONCLUSIONS

In this article, I have discussed different approaches to the problem of word sense division in WSD, with a focus on the way in which the LACELL project has handled overlaps among senses in pre-defined inventories. The proposal laid out by the LACELL group is to devise an algorithm capable of automatically detecting the maximum level of sense granularity at which disambiguation of a word occurrence is possible or relevant. Accordingly, the LACELL WSD system is not programmed to assign only one dictionary sense to each case. Where two or more senses are found to be equally probable in a particular context, the program selects then all, rather than attempting to disambiguate between/among them. That is, if the context is not sufficiently specific to discriminate between two or more senses, the program will capture this lack of discriminability, rather than analytically imposing an artificial process of fine-grained disambiguation which is not underpinned by objective properties of the text.

This decision has both positive and less positive consequences. On the positive side of the matter lies the possibility to cope with vagueness, which is an objective characteristic of many instances in natural language text. Sometimes the correct interpretation of a word is represented by a bunch of senses, a fact that is not accounted for by those WSD systems which aim to systematically project only one specific sense of a lexical entry onto each

occurrence of a word. The conjunction of several subsenses is a plausible strategy for resolving those cases in which the context is only weakly discriminant. However, this requires that the set of subsenses selected be sufficiently homogeneous to be subsumed under a single taxonomic category. Some contextual features are unable to discriminate at a coarse-grained level, for instance between senses that belong in separate conceptual domains, say, 'concrete-literal' (physical) and 'abstract-figurative'. This is the case of 'versatile collocates', i.e. lexical co-occurrences that are shared by literal and figurative senses of a word. In such cases, the contextual similarity among senses is a record of their phylogenetic relationship, i.e. of the fact that one of them has been originated from the other, but the senses in question need not form a sufficiently coherent category to be selected simultaneously.

The problem mentioned above suggests that WSD is a valuable testing ground for theoretical lexical semantics, and not merely a practical NLP application. At this point, it is appropriate to recall the words of Senseval-2 organizers stating that "[Senseval's] underlying mission is to develop our understanding of the lexicon and language in general" (Edmonds and Kilgarriff 2002; apud Ide and Wilks 2006: 48). Indeed, analyses of WSD performance can be used as empirical tests of the semantic theories which inspire the lexical resources employed. In turn, the results will beg questions that are relevant to further developments of theories and models in the field of lexical semantics. Thus, one interesting question raised by the analysis of results obtained with the LACELL WSD system concerns the nature of contextual similarity among senses. As a rule, contextual similarity between two or more uses of a word is closely aligned with semantic similarity, but it is also important to take into account the fact that the meaning of some collocates is susceptible to co-variation with the meaning of the node. Such collocates are less useful for disambiguating the sense of the node word than for keeping track of sense extension processes.

REFERENCES

Agirre, E. & Edmonds, P. (2006). Introduction. In E. Agirre & P. Edmonds (Eds.), *Word sense disambiguation: algorithms and applications*, Dordrecht: Springer, pp. 1-28.

Deignan, A. (2005). *Metaphor and corpus linguistics*. Amsterdam / Philadelphia: John Benjamins.

Fellbaum, C., Delfs, L., Wolff, S. & Palmer, M. (2005). Word meaning in dictionaries, corpora and the speaker's mind. In Barnbrook, G., P. Danielsson & M. Mahlberg (Eds.), *Meaningful texts. The extraction of semantic information from monolingual and multilingual corpora*. London / New York: Continuum, pp. 31-38.

Frisson, S. & Pickering, M. J. (1999). The processing of metonymy: evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25:6, 1366-1383.

Ide, N. & Véronis, J. (1998). Word sense disambiguation: the state of the art. *Computational Linguistics*, 24:1, 1-41.

Ide, N. & Wilks, Y. (2006). Making sense about sense. In E. Agirre & P. Edmonds (Eds.), *Word sense disambiguation: algorithms and applications*, Dordrecht: Springer, pp. 47-73.

Kilgarriff, A. (1997). 'I don't believe in word senses'. *Information Technology Research Institute Technical Report Series* (University of Brighton, *ITRI-97-12*).

Kilgarriff, A. (2001). Generative Lexicon meets corpus data: the case of non-standard word uses. *Information Technology Research Institute Technical Report Series* (University of Brighton, *ITRI-01-10*).

Kilgarriff, A. (2006). Word senses. In Agirre & P. Edmonds (Eds.), *Word sense disambiguation: algorithms and applications*, Dordrecht: Springer, pp. 29-46.

Klein, D. E. & Murphy, G. L. (2002). Paper has been my ruin: conceptual relations of polysemous senses. *Journal of Memory and Language*, 47: 548-570.

Pustejovski, J. (1995). *The Generative Lexicon*. Cambridge, MA / London: MIT Press.

Pylkkänen, L., Llinás, R. & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, 18:1, 97-109.

Resnik, P. & Yarowsky, D. (1999). Distinguishing systems and distinguishing senses: new evaluation methods for Word Sense Disambiguation. *Natural Language Engineering*, 5, 113-133.

Sánchez, A., Cantos, P. & Almela, M. (2007). Lexical constellations and the structure of meaning: a prototype application to WSD. In Gelbukh, A. (Ed.), *Computational Linguistics and Intelligent Text Processing, 8th International Conference, CICLing 2007, Mexico City*. Berlin: Springer, pp. 275-287.

Stathi, K. (2007): A corpus-based analysis of adjectival modification in German idioms. In: Fellbaum, C. (ed.), *Idioms and collocations. Corpus-based linguistic and lexicographic studies*. London / New York: Continuum.

Véronis, J. (2003). Sense tagging: does it make sense? In Wilson, A., P. Rayson & T. McEnery (Eds.), *Corpus Linguistics by the Lune. A Festschrift of Geoffrey Leech*. Frankfurt a. M.: Peter Lang, pp. 273-290.

