A Non-Distributional Approach to Polysemy Detection in WordNet

Tony Veale[1]

Abstract. While polysemy is a form of ambiguity that can complicate natural language processing, it is also a rich lexical resource that yields useful insights into the mapping between words and concepts. WordNet, a comprehensive lexical knowledge-base of English word meanings, is replete with instances of polysemy, but also contains many instances of homonymy, and fails to distinguish between both kinds of ambiguity. We propose in this paper an alternative to the distributional approach for recognizing polysemous sense-pairs in WordNet. Our approach does not rely on the systematicity of regular polysemy to identify the families of words that instantiate a particular metonymic pattern, but seeks instead local ontological evidence for each word, on a case by case basis, before polysemy is hypothesized. We argue that this approach is thus equally adept at recognizing instances of ad-hoc polysemy.

1 INTRODUCTION

Ambiguity is such a vexing problem in natural language processing [1] that it is easy to forget that, like cholesterol, lexical ambiguity comes in both a good form and a bad form. Homonymy, the bad form, is ambiguity arising from historical coincidences of language that do not follow any predictable conceptual patterns, and which generally serve no useful purpose beyond the generation of puns. The study of homonymy may illuminate in some small way the diachronic development of language but sheds no light at all on its conceptual underpinnings. Polysemy, on the other hand, arises when two or more related meanings are shoehorned into the same lexical form for reasons of linguistic economy or creativity. Polysemy is thus a good form of lexical ambiguity [2]. It's presence, if detected, can reveal the workings of a systematic conceptual trend at work, or a relational similarity between senses that has not been explicitly marked in the lexicon.

The key phrase here is "if detected". WordNet, a comprehensive ontologically-structured lexicon of English [3], is rich in instances of polysemy, but it is also home to many instances of homonymy and does not explicitly differentiate one from the other. This deficit can lead to false rationalizations and silly inferences. The fact that the word "bank" has both a river-side sense and a financial institution sense does not mean that the latter is to be found on the former. However, the fact that "bank" has a building sense and a financial institution sense does mean that the latter is to be found in the former. The building/financial-institution ambiguity of "bank" is an example of polysemy that instantiates a general conceptual tendency to conflate organizations with their locations (think "Whitehouse" and "Wall Street"), while the building/river-side ambiguity is simply a case of homonymy (each use of "bank" having a different etymological origin).

NLP systems must be able to distinguish polysemy from homonymy since each demands a different resolution mechanism. Unlike homonymy, where a single choice of senses must be made, polysemous words can be used in multiple different senses simultaneously [4,5]. For example, the sentence "the book was badly written but beautifully produced" requires two related senses of "book" to be co-active – "book" as a container of abstract content ("badly written") and as a physical artifact ("beautifully produced"). Since polysemy is also a (semi-)productive phenomenon, it allows an NLP system to dynamically plug the holes in its lexicon. Most dictionaries list "cod" and "haddock" as both a fish and a food, but few would bother to also list "shark" as a food. Yet "shark soup" requires a system to understand that sharks too are edible fish. The basis for this productivity lies in the existing polysemy patterns of language; sharks are fish, but "fish" can denote both a marine animal and the food derived from it.

In this paper we present an ontological basis for detecting polysemy patterns in WordNet. The approach differs from past work by working not at the general level of distributional analysis and word families as a whole, but at the specific level of individual words, seeking local ontological evidence for each instance of polysemy that it hypothesizes. This significantly reduces the possibility of falsely identifying homonymy as polysemy, while in many cases revealing, in relationally specific terms, the conceptual motivation of the polysemy.

2 PAST WORK

Polysemy is pervasive in the lexicon because it is, to a large extent, a productive phenomenon that arises from the conflation of general categories like Location, Person,

[1] Department of Computer Science, University College Dublin, Belfield, Dublin, Ireland.

Animal, Food, and so on. This productivity means that a particular polysemy pattern may be instantiated by a substantial family of different words. For instance, WordNet contains 344 words that can denote both a type of person and a language, reflecting our tendency to name languages after the peoples that speak them. Furthermore, the same polysemy pattern may be observed at different ontological levels – WordNet contains 158 words that denote both a grouping of American Indians and the language spoken by that grouping. This potential for polysemy to be regular and systematic has been expressed by [3] thus: "Polysemy of the word A with the meanings a_i and a_j is called regular if, in the language, there exists at least one other word B with the meanings b_i and b_j, which are semantically distinguished from each other in exactly the same way".

This formulation by Aprejan [3] is sufficiently algorithmic to yield a means of detecting polysemy via its systematic effect on the lexicon. This insight is applied to WordNet in [6,7], where the potential polysemy of two word senses is categorized in terms of their divergent positions in the WordNet ontology. For instance, "waltz" has a sense that lies under the hypernym {music} and another that lies under {dancing, dance, terpsichore}. Because this divergence is precisely mirrored by other words such as "Samba", "Rumba" and "Tango", it is possible to locate a sizeable family of words (23 in WordNet 1.6) that can denote both a kind of dancing and the kind of music that accompanies it. The larger the family of words that can be found to support a given divergence pattern, the more evidence there is to assume that the pattern captures a systematic tendency and that the ambiguity in each case is the result of polysemy rather than homonymy.

This distributional approach, prefigured by [4] and principally elaborated by [6,7], provides an excellent means of finding significant tendencies toward polysemy in the lexicon. But in exploiting the systematicity of regular polysemy, the distributional approach is prone to three forms of error. First, while it can reliably identify patterns like Animal/Food that give rise to word families like "turkey", "lamb", "chicken", "hen" and so on, it cannot reliably exclude homonyms from these families. For instance, "mate" the berry-tea drink is not derived from the animal sense of "mate", despite the fact that a significant 193 words instantiate the Animal/Food pattern. Secondly, the distributional approach ignores the fact that polysemy patterns are not transitive. For instance, WordNet defines several animal senses for "hen", the primary sense describing an adult female chicken and another, extended sense describing the female of certain aquatic animals such as lobsters and octopuses. However, the food sense of "hen" is a metonymic extension of the chicken sense only, and should not be used to denote the food obtained from lobsters and octopuses. Thirdly, some of the most interesting polysemy is, or at least appears to be, ad-hoc. WordNet only contains one word that denotes both a place of business and a person, "florist", yet this is a very useful polysemy pattern for an NLP system to comprehend. When one says "I went to the dentist", the intended meaning is more specific than "I went to the location of the dentist", which can allow odd interpretations like "I went to the restaurant where the dentist was eating". Tight metonymic connections like that between places of business and businessmen are worth extracting from WordNet even if they happen to be under-represented and thus appear ad-hoc.

Systematic tendencies toward polysemy can be represented in an ontology like WordNet quite effectively, by connecting the general categories involved with lexical rules. For instance, a connection between {animal} and {food} can be used to imply that any word that has both an animal sense and a food sense is a product of this polysemous tendency. This is the approach employed by the WordNet *cousins* mechanism [8], so-called because the word-families that are created as a result – such as "turkey", "lamb", and "chicken" for the {animal}/{food} connection – are deemed to be lexical cousins. The problem, of course, is that such rules have many exceptions, and because these exceptions do not obey easy generalizations, they must be exhaustively listed. WordNet is thus forced to list an average of 15 exceptions for every cousin rule. The extent of this list represents a warning for all distributional approaches, suggesting that homonymy can be a very compelling null hypothesis when a word has multiple senses.

3 A LOCAL, EVIDENTIAL APPROACH

Polysemy is a form of lexical ambiguity where different senses are psychologically related. So to recognize polysemy between word senses in WordNet without exploiting a distributional analysis, it is worth considering how these specific senses are connected, either ontologically, or via their glosses. We take as an example the word "olive", which has five noun senses in WordNet 1.6 (note: isa+ indicates hypernymy via transitive closure, and superscripts indicate the junctures at which one sense suggests a relationship with another):

olive.1 *gloss:* a yellow^3 green color of low brightness and saturation
isa: {chromatic_color, spectral_color}

olive.2 *gloss:* one-seeded fruit^5 of the European olive tree^4 usually pickled …
isa: {relish}

olive.3 *gloss:* hard yellow^1 variegated wood of an olive tree^4
isa: {wood^4} *which isa* {plant^4 material}

olive.4 *gloss:* evergreen tree cultivated in the Mediterranean region …
isa: {olive tree} *which isa* {fruit^5 tree}
which isa$^+$ {wood^{3}y plant}

olive.5 *gloss:* small ovoid fruit of the European olive tree^4
isa: {fruit^4} *which isa*$^+$ {plant^4 organ}

First, note how three senses of "olive", 2,3 and 5, directly reference the concept {olive tree} in their glosses. We can assume then that these senses are all derivations of sense 4, which specifies {olive tree} as a hypernym. Now consider

how some senses seem to relate to another sense via some kind of ontological *frame:slot* filling. Sense 5 is a {fruit} and sense 4 is a {fruit tree}, so it seems likely that sense 5 denotes the fruit of the tree denoted by sense 4. Sense 3 denotes a kind of {wood}, while sense 4 denotes a kind of {woody plant}, so again, it seems likely that sense 3 denotes the wood derived from the tree denoted by sense 4 (simple morphology is needed to make this connection). However, sense 1, the color sense, seems problematic until we realize that it can be considered a hyponym of the chromatic color {yellow, yellowness}. WordNet does not state this connection, but it can be inferred since sense 1 is also a chromatic color that mentions "yellow" in its gloss. Thus, sense 3, the wood sense, can be seen to refer to olive.1 in its gloss by its use of the term "yellow".

In effect then, the information needed to recognize the five noun senses of "olive" as mutually polysemous can be found in the definitions of these senses themselves. It is not necessary to look outside the senses to find distributional evidence elsewhere for what is already stated, in a somewhat implicit form, as part of the senses themselves. This is the essence of the evidence-based approach to polysemy detection: a variety of ontologically-motivated connection strategies are used to identify the implicit relationships between senses to support the hypothesis that these senses form a polysemous bond. When no evidence is found, we err on the side of caution and assume homonymy. In the case of "olive" above, the sense-pairings were produced by applying the following two strategies:

Explicit Ontological Bridging: a sense pair $<\omega_1, \omega_2>$ for a word ω can be bridged if ω_1 has a hypernym that can be lexicalized as M-H and ω_2 has a hypernym that can be lexicalized as M, the rationale being that ω_2 is the M of ω_1 and ω_1 is the H of ω_2. E.g., the word "basketball" has two WordNet senses, one a transitive hyponym of {game}, the other a hyponym of {game equipment}. In this case then, M = *game* and H = *equipment*. The second sense thus denotes the equipment used in the activity of the first sense.

Cross-Reference: if $<\omega_1, \omega_2>$ is a sense pair for a word ω and the WordNet gloss for ω_2 explicitly mentions a hypernym of ω_1, then ω_2 can be seen as a conceptual extension of ω_1. For instance, WordNet contains several senses of the word "charcoal", one of which is a hyponym of {drawing}, another of which refers to "drawing" in its gloss. The latter sense, a hyponym of {writing implement}, can thus be seen as making a reference to the former. Cross-reference is a powerful connection strategy, and is all the more powerful for considering the glosses of sense hypernyms as well. For instance, WordNet defines "angler" as both a {fisherman, fisher} and as a type of acanthopterygian {fish} that lures other fish as its prey. Since the gloss for {fisherman, fisher} makes reference to "fish", the polysemous link between both senses of angler can be detected.

No one strategy is powerful enough to recognize all inter-sense relationships. We see this approach not as an essentialist account of polysemy, but as an engineering approach to detecting the tell-tale connections that exist between sense descriptions in a hand-built ontology like WordNet. We thus engage a wide variety of different strategies, each capturing a different intuition about sense definitions and the way they reflect linguistic knowledge.

Hierarchical Reinforcement: if $<\alpha_1, \alpha_2>$ and $<\beta_1, \beta_2>$ are sense pairs for two words α and β where α_1 is a hypernym of β_1 and α_2 is a hypernym of β_2, then $<\alpha_1, \alpha_2>$ reinforces the belief that $<\beta_1, \beta_2>$ is polysemous, and vice versa. For example, "herb" can denote either a plant or a foodstuff in WordNet, while the words "sage", "dill", "coriander", "cilantro" and twenty others can denote a subclass of either of these senses. If words like "herb" were truly homonymous, we would not expect their ambiguity, essentially an accident of language, to be mirrored at their subclass level. Hierarchical reinforcement is essentially a special case of the distributional approach, applying Apresjan's intuition about systematicity to the local context of a word. Nonetheless, it is also an evidential strategy, seeking word-specific evidence before polysemy is hypothesized.

Blend Recruitment: a sense pair $<\omega_1, \omega_2>$ for a word ω can be related by this strategy if ω_1 has a hypernym that can be lexicalized as M-H_1, and ω_2 has a hypernym that can be lexicalized as H_2, and M-H_2 is a known hyponym of the same sense of H_2. We refer to M-H_2 as the recruited blend. For example, WordNet associates with "commander" a sense which is a hyponym of {military officer} and another which is a hyponym of {leader}. The blended middle ground between these senses is clearly the concept {military leader}, which WordNet defines as a hyponym of {leader}.

Morphosemantic Linking: a sense pair $<\omega_1, \omega_2>$ for a word ω can be related by this strategy if ω_1 has a hypernym that can be lexicalized as H_1, and ω_2 has a hypernym that can be lexicalized as H_2, and H_1 is morphologically derived from H_2 or vice versa. For instance, WordNet attributes a sense to "gossip" that is a hyponym of {communicator}, and another that is a hyponym of {communication} via {chat, confab, confabulation}. This suggests that both senses engage in a *communicator:communication* relationship. Morphosemantic linkages between synsets like {communication} and {communicator} are now provided as standard as part of WordNet 2.0, obviating the need for morphology rules to achieve this linkage.

4 APPLICATIONS OF POLYSEMY

It is perhaps just as well that WordNet does not explicitly mark instances of polysemy, since an understanding of how to dynamically recognize polysemous word usage is far more useful to an NLP system than a static list of conventional ambiguities. For instance, consider the problem of lexical coherence and how it might be addressed using the technique of lexical chaining [9].

(a) When she sings an aria, the song is soon immortalized.

Chaining allows an NLP system to recognize the coherence between the two clauses of (a), linking the word "song" in the second clause to the word "aria" in the first.

(b) When she sings a song, the performance is lauded.
(c) When she sings a lovesong, her performance is tearful.

In (b) it is possible to construct a chain linking "sings" to "song" and linking "song" to "performance", because one sense of "song" in WordNet is defined as a hyponym of {performance, doing, execution}. However, though this chain seems equally possible in (c), WordNet prevents its construction by providing only one sense of "lovesong", as a hyponym of {song} *qua* musical object rather than as a hyponym of song *qua* performance. This seems a rather false economy since other hyponyms of {song}, such {lullaby} and {aria}, reflect this music/performance duality. However, since {lovesong} is defined as a hyponym of {song} *qua* musical object, and since this sense is related via polysemy to {song} *qua* performance, an NLP system can infer that "lovesong" too should be given, dynamically, an extended sense of {lovesong} *qua* performance. This kind of sense completion – a highly constrained form of sense extension that exploits the systematicity of existing sense structures – effectively uses a knowledge of polysemy patterns to dynamically plug any sense gaps in the lexicon.

Dynamic sense completion also provides an extra source of flexibility in applying a preferential semantics to verb role filling [10]. For instance, what selectional restriction should the verb "perform" place on its patient in (d)?

(d) When she performs a lovesong, the audience swoons.

At first one might be tempted to suggest {performance, doing, execution} as a restriction, allowing any word that denotes a hyponym of this category, such as "music" or "song", to fill the patient role. This would exclude "lovesong" of course, since WordNet does not define it as a {performance, doing, execution}. However, sense completion would allow a performance sense to be created dynamically for the purpose of role filling.

Strictly speaking however, one does not perform a performance – a performance is the event that arises as a result of executing a *performable* action. Though WordNet does not provide a neat ontological definition of "performable" action, it can be defined neatly enough in terms of polysemy: a performable action is one that can be connected to the hyponym {performance, doing, execution} either directly or via sense completion. Thus, one can talk of "performing a jig" even though WordNet does not define {folk dancing} as a kind of performance.

These examples show how an understanding of polysemy can increase the level of sense integration in a lexical ontology. Consider the word "monopoly", which WordNet defines both as a kind of {market, marketplace} and as a state of {dominance, control, ascendancy}. Neither of these senses alone is effective in capturing the true sense of monopoly – a state of dominance applied to a market – and one suspects that this unnecessary splintering of senses is done only because WordNet lacks a sufficient ontological description language to formulate such a composition of taxonomically different ideas. A similar fragmentation is evident in the sense structure of "disaster", which is divided into three senses, an action, a state and an event. Only by recognizing these senses as polysemous can the full concept of disaster be understood.

Once these sense splinters are understood as part of a coherent whole, we can exploit WordNet's inherent fragmentation for more useful purposes. For instance, if a concept like monopoly is understood as an integration of two senses, we can modulate one of the senses in a well-defined way to generate metonyms and metaphors that can be used for query expansion in creative information retrieval [11]. Thus, since a monopoly is defined as a state of dominance in a business market ({dominance, control, ascendancy} + {market, marketplace}), we can replace the dominance facet with more specific hyponyms, generating metaphors like "market mastery" , "market tyranny" and "market despotism". In this way polysemy provides a means of making creative leaps around the space of possible lexicalizations [12]. Ultimately this should not be too surprising since polysemy is, after all, an artifact of linguistic creativity.

5 EVALUATION

To test the coverage and precision of various polysemy detection strategies, we can compare the set of specific polysemy predictions that they generate against those generated by the WordNet cousin rules [8].

Table 1. Estimations of coverage and precision as measured against WN cousins.

Strategy	Coverage	Precision
Ontological Bridging	15%	95%
Cross-Reference	76%	89%
Hierarchical Reinforcement	11%	76%
Blend Recruitment	3%	89%
Morphosemantic Linking	3%	93%
All 25 Strategies combined	98%	85%

Of course, the cousin rules provide only a partial account of the polysemy inherent in WordNet, connecting senses in just 20% of the ambiguous nouns in WordNet 1.6. In contrast, the 25 evidential strategies we have implemented so far (section 3 describes the most significant ones) succeed in connecting at least one pair of senses in 70% of all of WordNet's multi-sense nouns. Though the specific

breakdown between polysemous and homonymous words in WordNet is unknown, we can nevertheless estimate the coverage of these strategies by calculating the percentage of cousin predictions that each strategy manages to independently replicate. Furthermore, because the cousin rules have a manually constructed exception list, this allows us to also estimate the precision of each strategy, by calculating the percentage of hypothesized instances of polysemy that do not correspond to cousin exceptions. The results of these estimations are shown in Table 1, reported on a per-strategy basis and for all 25 strategies combined.

The results suggest that WordNet contains enough information in its ontological structure and in its glosses to allow a non-distributional, word-by-word analysis of polysemy to succeed. The strategy that offers the most coverage is *cross-reference*, and the strategies that offer the least are *morphosemantic linking* and *blend recruitment*. The latter two are ontological while the former makes use of the shallow knowledge that resides in the sense glosses. However, though the purely ontological strategies offer less coverage, they yield greater structural insights into the conceptual roots of polysemy.

For instance, the strategy of *ontological bridging* can explain instances of polysemy in specific relational terms. The fruit sense of "olive" denotes the fruit of the olive tree, while the tree sense denotes the tree from which the olive fruit is obtained, so these senses can be automatically connected via the relations *fruit-of* and *tree-of* respectively. Likewise, the *morphosemantic linking* strategy not only suggests that two senses of "gossip" are related, but that the relationship is one of *communicator-of* to *communication-of*. Furthermore, the *blend recruitment* strategy suggests a conceptual middle ground between two senses that are related by polysemy. Thus, the moral sense of "war", a hyponym of {struggle, conflict, battle}, and the military sense, a hyponym of {military action}, can be reconciled via the blended concept {military campaign}. The internet sense of "web", a hyponym of {computer network}, can likewise be reconciled with the social sense, a hyponym of {system}, via the blend {computer system}. While not every strategy can offer such insight into the conceptual basis of polysemy, the results of these strategies at least are certainly suggestive.

6 CONCLUSIONS

Polysemy is not a weakness of WordNet but a strength, one that can open a richly illuminating window on lexico-conceptual structure if approached from the right perspective. The evidence-based approach described here offers an essentially bottom-up perspective on polysemy detection, while the distributional approach offers what is essentially a top-down perspective. However, both perspectives can complement each other, to help mitigate the weaknesses of both. The evidence-based approach can be stymied by opaque glosses and counter-intuitive ontological definitions, while the distributional approach is weak in dealing with ad-hoc polysemy and individual instances of homonymy that happen to only coincidently conform to systematic patterns of polysemy.

In an ideal lexical ontology, polysemy detection strategies would not rely on a shallow knowledge source like flat text glosses (as in *cross-reference*), but would derive all of their guidance from the logical structure of the ontology. It is instructive to imagine what such an ontology might look like, if the conceptual motivation for each instance of polysemy in a given language were to be somehow made explicit in its underlying taxonomic structure. As we look for stronger strategies to replace *cross-reference*, it is perhaps worth considering how we might change WordNet itself, both through automatic and manual activities, so that it may truly explain, rather than merely report, polysemy.

REFERENCES

[1] Y. Ravin, C. Leacock, Polysemy: An Overview. *Polysemy: Theoretical and Computational Approaches*, eds. Ravin, Y, Leacock, C. Oxford University Press, Oxford (2000).

[2] G. A. Miller, WordNet: A Lexical Database for English. *Communications of the ACM*, Vol. 38 No. 11 (1995)

[3] J. D. Apresjan, Regular Polysemy. *Linguistics*, 142:5-32.

[4] J. Pustejovsky, The Generative Lexicon. *Computational Linguistics* 17(4).

[5] D. A. Cruse, Aspects of the Micro-structure of Word Meanings. *Polysemy: Theoretical and Computational Approaches*, eds. Ravin, Y, Leacock, C. Oxford University Press, Oxford (2000).

[6] W. Peters, I. Peters, Extracting Regular Polysemy Patterns in WordNet. *Technical Report*, University of Sheffield, UK. (2000)

[7] W. Peters, I. Peters, Lexicalized Systematic Polysemy in WordNet. *Proceedings of the 2nd international conference on Language Resources and Evaluation.* Athens. (2000)

[8] WordNet documentation. *www.princeton.edu/~wn/* (2003)

[9] G. Hirst, D. St-Onge, Lexical Chains as a Representation of Context for the Detection of Malapropisms, *WordNet: An Electronic Lexical Database*, C. Fellbaum ed., MIT Press, Cambridge MA, (1998).

[10] Wilks, Y. A preferential, pattern-Seeking, semantics for natural language inference, *Artificial Intelligence* **6**, pp 53-74, (1974).

[11] T. Veale, Creative Information Processing. *Proceedings of CICLing 2004*, Seoul, Korea (2004).

[12] T. Veale, Pathways to Creativity in Lexical Ontologies. *Proceedings of The 2nd Global WordNet Conference*, Brno, Czech Republic (2004).