Kenneth C Litkowski CL Research 9208 Gue Road Damascus, MD 20872 ken@clres com http://www.clres.com

Abstract

The mapping from WordNet to Hector senses in Senseval provides a "gold standard" against which to judge our ability to compare lexical resources The "gold standard" is provided through a word overlap analysis (with and without a stop list) for this mapping, achieving at most a 36 percent correct mapping (inflated by 9 percent from "empty" assignments) An alternative componential analysis of the definitions, using syntactic, collocational, and semantic component and relation identification (through the use of defining patterns integrated seamlessly into the parsing dictionary), provides an almost 41 percent correct mapping, with an additional 4 percent by recognizing semantic components not used in the Senseval mapping Definition sets of the Senseval words from three published dictionaries and Dorr's lexical knowledge base were added to WordNet and the Hector database to examine the nature of the mapping process between definition sets of more and less scope. The techniques described here constitute only an initial implementation of the componential analysis approach and suggests that considerable further improvements can be achieved

Introduction

The difficulty of comparing lexical resources, long a significant challenge in computational linguistics (Atkins, 1991), came to the fore in the recent Senseval competition (Kilgarriff, 1998), when some systems that relied heavily on the WordNet (Miller, et al, 1990) sense inventory were faced with the necessity of using another sense inventory (Hector) A hasty solution to the problem was the development of a map between the two inventories, but some participants expressed concerns that use of this map may have degraded their performance to an unknown degree

Although there were disclaimers about the WordNet-Hector map, it nonetheless stands as a usable gold standard for efforts to compare lexical resources Moreover, we have a usable baseline (a word overlap method suggested in (Lesk, 1986)) against which to compare whether we are able to make improvements in the mapping (since this method has been shown to perform not as well as expected (Krovetz, 1992))

We first describe the lexical resources used in the study (Hector, WordNet, other dictionaries, and a lexical knowledge base), first characterizing them in

terms of polysemy and the types of lexical information each contains (syntactic properties and features, semantic components and relations, and collocational properties) We then present results of performing the word overlap analysis of the 18 verbs used in Senseval, analyzing the definitions in WordNet and Hector We then expand our analysis to include other dictionaries We describe our methods of analysis, particularly the methods of parsing definitions and identifying semantic relations (semrels) based on defining patterns. essentially taking first steps in implementing the program described by Atkins and focusing on the use of "meaning" full information rather than statistical information We identify the results that have been achieved thus far and outline further steps that may add more "meaning" to the analysis 1

¹All analyses described in this paper were performed automatically using functionality incorporated in DIMAP (Dictionary Maintenance Programs) (available for immediate download at (CL Research, 1999a)) This includes automatic extraction of WordNet information for the selected words (integrated in DIMAP) Hector definitions were uploaded into DIMAP dictionaries after use of a conversion program Definitions for other

The Lexical Resources

This analysis focuses on the main verb senses used in Senseval (not idioms and phrases), specifically the following

AMAZE, BAND, BET, BOTHER, BURY, CALCULATE, CONSUME, DERIVE, FLOAT, HURDLE, INVADE, PROMISE, SACK, SANCTION, SCRAP, SEIZE, SHAKE, SLIGHT

The Hector database used in Senseval consists of a tree of senses, each of which contains definitions, syntactic properties, example usages, and "clues" (collocational information about the syntactic and semantic environment in which a word appears in the specific sense) The WordNet database contains synonyms (synsets), perhaps a definition or example usages (gloss), some syntactic information (verb frames), hypernyms, hyponyms, and some other semrels (ENTAILS, CAUSES)

To extend our analysis in order to look at other issues of lexical resource comparison, we have included the definitions or lexical information from the following additional sources

- Webster's 3rd New International Dictionary (W3)
- Oxford Advanced Learners Dictionary (OALD)
- American Heritage Dictionary (AHD)
- Dorr's Lexical Knowledge Base (Dorr)

We used only the definitions from W3, OALD, and AHD (which also contain sample usages and some collocational information in the form of usage notes, not used at the present time) Dorr's database contains thematic grids which characterize the thematic roles of obligatory and optional semantic components, frequently identifying accompanying prepositions (Olsen, et al, 1998)

The following table identifies the number of senses and average overall polysemy for each of these resources

Word	Hector	WordNet	W3	AHD	OALD	Dorr
amaze	1	2	4	2	1	2
band	3	1	11	4	2	4
bet	4	2	5	5	1	3
bother	7	6	9	7	4	4
bury	12	6	14	5	8	1
calculate	5	5	10	9	3	1
consume	6	6	8	8	3	1
derive	6	5	15	5	3	2
float	16	4	41	14	10	5
hurdle	2	1	4	3	1	0
invade	6	2	10	5	3	1
promise	5	4	7	4	3	2
sack	4	4	6	3	2	0
sanction	2	2	5	2	1	1
scrap	3	1	3	3	1	0
seize	11	6	21	13	7	1
shake	8	8	37	17	7	12
slight	1	1	6	3	1	0
Average Polysemy	57	37	12 0	6 2	34	2 2

Word Overlap Analysis

We first establish a baseline for automatic replication of the lexicographer's mapping from WordNet 1 6 to Hector, using a simple word overlap analysis similar to (Lesk, 1986) The lexicographer mapped the 66 WordNet senses (each synset in which a test occurred) into 102 Hector senses A total of 86 assignments were made, 9 WordNet senses were given no assignments, 40 received exactly one, and 17 senses received 2 or 3 assignments The WordNet senses contained 348 words (about half of which were common words appearing on our stop list, which contained 165 words, mostly prepositions, pronouns, and conjunctions) The Hector senses selected in the word overlap analysis contained about 960 words (all Hector senses contained 1878 words)

We performed a strict word overlap analysis (with and without a stop list) between the definitions in WordNet and the Hector senses, that is, we did not attempt to identify root forms of inflected words We took each word in a WordNet sense and determined whether it appeared in a Hector sense, we selected a Hector sense based on the highest percentage of words over all Hector senses An

dictionaries were entered by hand

empty selection was made if all the words in the WordNet sense did not appear in any Hector sense, only content words were considered when the stop list was used

For example, for bet, WordNet sense 2 (stake (money) on the outcome of an issue) mapped into Hector sense 4 ((of a person) to risk (a sum of money or property) in this way) In this case, there was an overlap on two words (money, of) in the Hector definition (0 13 of its 15 words) without the stop list When the stop list was invoked, there was an overlap of only one word (money, 0 07 of the Hector definition) In this case, the lexicographer had made three assignments (Hector senses 2, 3, and 4), our scoring method treated this as only 1 out of 3 correct (not using the relaxed method employed in Senseval of treating this as completely correct)

Without the stop list, our selections matched the lexicographer's in 28 of 86 cases (32 6%), using the stop list, we were successful in 31 of 86 cases (36 1%) The improvement arising when the stop list was used is deceptive, where 8 cases were due to empty assignments (so that only 23 cases, 26 7%, were due to matching content words) Overall, only 41 content words were involved in these 23 successes when the stop list was used, an average of 1 8 content words

To summarize the word overlap analysis (1) despite a richer set of definitions in Hector, 9 of 66 WordNet senses (13 6%) could not be assigned, (2) despite the greater detail in Hector senses compared to WordNet senses (2 8 times as many words), only 1 8 content words participated in the assignments, and (3) therefore, the defining vocabulary between these two definition sets seems to be somewhat divergent Although it might appear as if the word overlap analysis does not perform well, this is not the case The analysis provides a broad overview of the definition comparison process between two definition sets and frames a deeper analysis of the differences Moreover, it appears that the accuracy of a "gold standard" mapping is not crucially important The quality of the mapping may help frame the subsequent analysis more precisely, but it seems sufficient that any reasonable mapping will suffice This will be discussed further after presenting the results of the componential analysis of the definitions

Meaning-Full Analysis of Definitions

The deeper analysis of the mapping between two definition sets relies primarily on two major steps (1) parsing definitions and using defining patterns to identify semrels present in the definitions and (2) relaxing values to these relations by allowing "synonymic" substitution (using WordNet) Thus, for example, if we identify hypernyms or instruments from parsing a definition, we would say that the definitions are "equal" not just if the hypernym or instrument is the same word, but also if the hypernyms or instruments are members of the same synset

This approach is based on the finding (Litkowski, 1978) that a dictionary induces a semantic network where nodes represent "concepts" that may be lexicalized and verbalized in more than one way This finding implies, in general, the absence of true synonyms, and instead the kind of "concept" embodied in WordNet synsets (with several lexical items and phraseologies) A similar approach, parsing definitions and relaxing semrel values, was followed in (Dolan, 1994) for clustering related senses within a single dictionary

The ideal toward which this approach strives is a complete identification of the meaning components included in a definition The meaning components can include syntactic features and characteristics (including subcategorization patterns), semantic components (realized through identification of semrels), selectional restrictions, and collocational specifications

The first stage of the analysis parses the definitions (CL Research, 1999b, Litkowski, to appear) and uses the parse results to extract (via defining patterns) semrels Since definitions have many idiosyncrasies (that do not follow ordinary text), an important first step in this stage is preprocessing the definition text to put it into a sentence frame that facilitates the extraction of semrels 2

²Note that the stop list is not applicable to the definition parsing The parser is a full-scale sentence parser, where prepositions and other words on the stop list are necessary for successful parsing Moreover, inclusion of the prepositions is crucial to the method, since they are the bearers of much semrel information

The extraction of semrels examines the parse results, 1 e, a tree whose intermediate nodes represent non-terminals and whose leaves represent the lexical items that comprise the definitions, where any node may also include annotations such as characterizations of number and tense For all noun or verb definitions, this includes identification of the head noun (with recognition of "empty" heads) or verb, for verbs, we signal whether the definition contained any selectional restrictions (that is, particular parenthesized expressions) for the subject and object We then examine prepositional phrases in the definition and determine whether we have a "defining pattern" for the preposition which we can use as indicative of a particular semirel We also identify adverbs in the parse tree and look these up in WordNet to identify an adjective synset from which they are derived (if one is given)

The defining patterns are actually part of the dictionary used by the parser That is, we do not have to develop specific routines to look for specific patterns A defining pattern is a regular expression that articulates a syntactic pattern to be matched Thus, to recognize a "manner" semrel, we have the following entry for "in"

in(dpat((~ rep01(det(0)) adj manner(0)
sr(manner))))

This allows us to recognize "in" as possibly giving rise to a "manner" component, where we recognize "in" (the tilde, which allows us to specify particular elements before the "in" as well), with a noun phrase that consists of 0 or 1 determiner, an adjective, and the literal "manner" The '0' after the determiner and the literal indicate that these words are not copied into the value for a "manner" role, so that the value to the "manner" semrel becomes only the adjective that is recognized

The second stage of the analysis uses the populated lexical database to compare senses and make the selections This process follows the general methodology used in Senseval (Litkowski, to appear) Specifically, in the definition comparison, we first examine exclusion criteria to rule out specific mappings These criteria include syntactic properties (e g, a verb sense that is only transitive cannot map into one that is only intransitive) and collocational properties (e g, a sense that is used with a particle cannot map into one that uses a different particle) At the present time, these are used only minimally We next score each viable sense based on its semrels We increment the score if the senses have a common hypernym or if a sense's hypernyms belong to the same synset as the other sense's hypernyms If a particular sense contains a large number of synonyms (that is, no differentiae on the hypernym) and they overlap considerably in the synsets they evoke, the score can be increased substantially Currently, we add 5 points for each match ³

We increment the score based on common semrels In this initial implementation, we have defining patterns (usually quite minimal) for recognizing instrument, means, location, purpose, source, manner, has-constituents, has-members, is-part-of, locale, and goal ⁴ We increment the score by 2 points when we have a common semrel and then by another 5 points when the value is identical or in the same synset

After all possible increments to the scores have been made, we then select the sense(s) with the highest score Finally, we compare our selection with that of the gold standard to assess our mapping over all senses

Another way in which our methodology follows the Senseval process is that it proceeds incrementally Thus, it is not necessary to have a "final" perfect parse and mapping routine We can make continual refinements at any stage of the process and examine the overall effect As in Senseval, we may make changes to deal with a particular phenomenon with the result that overall performance declines, but with a sounder basis for making subsequent improvements

Results of Componential Analysis

The "gold standard" analysis involves mapping 66 WordNet senses with 348 words into 102 Hector senses with 1878 words Using the method described above, we obtained 35 out of 86 correct

³At the present time, we use WordNet to identify semrels We envision using the full semantic network created by parsing all a dictionary's definitions This would include a richer set of semrels than currently included in WordNet

⁴The defining patterns are developed by hand We have only just begun this effort, so the current set is somewhat impoverished

mappings (40 7%), a slight improvement over the 31 correct assignments using the stop-list word overlap technique However, as mentioned above, the stop-list technique had achieved 8 of its successes by matching null assignments Considered on this basis, it seems that the componential analysis technique provides substantial improvement In addition, our technique "erred" on 4 cases by making assignments where none were made by the lexicographer We suggest that these cases do contain some common elements of meaning and may conceivably not be construed as errors

Perhaps more importantly, the componential analysis method exploits considerably more information than the word overlap methods Whereas the stop-list word overlap mapping was based on only 41 content words, the componential approach (in the selected mappings) had 228 hits in developing its scores, with only a small number of defining patterns

Comparison of Dictionaries

We next examined the nature of the interrelations between pairs of dictionaries without use of a "gold standard" to assess the process of mapping For this purpose, we mapped in both directions between the pairs {WordNet, Hector}, {W3, OALD}, and {W3, AHD We examine Dorr's lexical knowledge base for the implications it may have in the mapping process

Neither WordNet nor Hector are properly viewed as dictionaries, since there was no intention to publish them as such WordNet "glosses" are generally smaller (5 3 words per sense) compared to Hector (18 4 words per sense), which contains many words specifying selectional restrictions on the subject and object of the verbs Hector was used primarily for a large-scale sense tagging project The three formal dictionaries were subject to rigorous publishing and style standards The average number of words per sense were 8 7 (OALD), 7 1 (AHD), and 9 9 (W3), with an average of 3 4, 6 2, and 12 0 senses per word

Each table shows the average number of senses being mapped, the average number of assignments in the target dictionary, the average number of senses for which no assignment could be made, the average number of multiple assignments per word, and the average score of the assignments that were made The mapping from WordNet to Hector had relatively few empty mappings, senses for which it was not possible to make an assignment These are the cases where it appears that the dictionaries do not overlap and thus provide a tentative indication of where two dictionaries may have different coverage The cases of multiple assignments indicate the degree of ambiguity in the mapping The average in both directions between Hector and WordNet were dominated by the inability to obtain good discrimination for the word "seize" Thus, this method identifies individual words where the discriminative ability needs to be further refined

WordNet - Hector						
	Senses	Assignments	Empty	Multiple	Scores	
WN-Hector	37	47	06	17	119	
Hector-WN	57	64	14	22	113	

These points are further emphasized in the mapping between W3 and OALD, where the disparity between the empty and multiple assignments indicate that we are mapping between dictionaries quite disparate This tends to be the case not only for the entire set of words, but also is evident for individual words where there is a considerable disparity in the number of senses, which then dominate the overall disparity Thus, for example, W3 has 41 definitions for "float", while OALD has 10 We tend to be unable to find the specific sense in going from W3 to OALD, because it is likely that we have many more specific definitions that are not present In the other direction, we are likely to have considerable ambiguity and multiple assignments



	Senses	Assignments	Empty	Multiple	Scores
W3-OALD	12 0	78	60	18	99
OALD-W3	34	60	07	32	86

Between W3 and AHD, there is less overall disparity between the definition sets, although since W3 is unabridged, we still have a relatively high number of senses in W3 that do not appear to be present in AHD Finally, it should be noted that the scores for the published dictionaries tend to be a little lower than for WordNet and Hector This reflects the likelihood that we have not extracted as much information as we did in parsing and analyzing the definition sets used in Senseval

W3 - AHD							
	Senses	Assignments	Empty	Multiple	Scores		
W3-AHD	12 0	115	40	36	90		
AHD-W3	62	91	12	4 1	91		

We next considered Dorr's lexical database We first transformed her theta grids into syntactic specifications (transitive or intransitive) and identification of semrels (e g, where she identified an **instr** component, we added such a semrel to the DIMAP sense) We were able to identify a mapping from WordNet to her senses for two words ("float" and "shake") for which Dorr has several entries However, since she has considerably more semantic components than we are currently able to recognize, we did not pursue this avenue any further at this time

More important than just mapping between two words, Dorr's data indicates the possibility of further exploitation of a richer set of semantic components Specifically, as reported in (Olsen, et al , 1998), in describing procedures for automatically acquiring thematic grids for Mandarin Chinese, it was noted that "verbs that incorporate thematic elements in their meaning would not allow that element to appear in the complement structure" Thus, by using Dorr's thematic grids when verb are parsed in definitions, it is possible to identify where particular semantic components are lexicalized and which others are transmitted through to the thematic grid (complement or subcategorization pattern) for the definiendum

The transmission of semantic components to the thematic grid is also reflected overtly in many

definitions For example, shake has one definition, "to bring to a specified condition by or as if by repeated quick jerky movements" We would thus expect that the thematic grid for this definition should include a "goal" And, indeed, Dorr's database has two senses which require a "goal" as part of their thematic grid Similarly, for many definitions in the sample set, we identified a source defining pattern based on the word "from," frequently, the object of the preposition was the word "source" itself, indicating that the subcategorizationproperties of the definiendum should include a source component

Discussion

While the improvement in mapping by using the componential analysis technique (over the word overlap methods) is modest, we consider these results quite significant in view of the very small number of defining patterns we have implemented Most of the improvement stems from the word substitution principle described earlier (as evidenced by the preponderance of 5 point scores) This technique also provides a mechanism for bringing back the stop words, viz, the prepositions, which are the carriers of information about semrels (the 2 point scores)

The more general conclusion (from the word substitution) is that the success arises from no longer considering a definition in isolation. The proper context for a word and its definitions consists not just of the words that make up the definition, but also the total semantic network represented by the dictionary

We have achieved our results by exploiting only a small part of that network We have moved only a few steps into that network beyond the individual words and their definitions We would expect that further expansion, first by the addition of further and improved semrel defining patterns, and second, through the identification of more primitive semantic components, will add considerably to our ability to map between lexical resources We also expect improvements from consideration of other techniques, such as attempts at ontology alignment (Hovy, 1998)

Although the definition analysis provided here was performed on definitions within a single language, the various meaning components correspond to those used in an Interlingua The use of the extinction method (developed in order to characterize verbs in another language, Chinese) can fruitfully be applied here as well

Two further observations about this process can be made The first is that reliance on a wellestablished semantic network such as WordNet is not necessary The componential analysis method relies on the local neighborhood of words in the definitions, not on the completeness of the network Indeed, the network itself can be bootstrapped based on the parsing results The method can work with any semantic network or ontology and may be used to refine or flesh out the network or ontology

The second observation is that it is not necessary to have a well-established "gold standard" Any mapping will do All that is necessary is for any investigator (lexicographer or not) to create a judgmental mapping The methods employed here can then quantify this mapping based on a word overlap analysis and then further examine it based on the componential analysis The componential analysis method can then be used to examine underlying subtleties and nuances in the definitions, which a lexicographer or analyst can then examine in further detail to assess the mapping

Future Work

This work has marked the first time that all the necessary infrastructure has been combined in a rudimentary form Because of its rudimentary status, the opportunities for improvement are quite extensive In addition, there are many opportunities for using the techniques described here in further NLP applications

First, the techniques described here have immediate applicability as part of a lexicographer's workstation When definitions are parsed and semrels are identified, the resulting data structures can be applied against a corpus of instances for particular words (as in Senseval) for improving word-sense disambiguation The techniques will also permit comparing an entry with itself to determine the interrelationships among its definitions and of comparing the definitions of two "synonyms" to determine the amount of overlap between them on a definition by definition basis

Although the analysis here has focused on the parsing of definitions, the development of defining patterns clearly extends to generalized text parsing Since the defining patterns have been incorporated into the same dictionary used for parsing free text. the patterns can be used directly to identify the presence of particular semrels among sentential constituents We are working to integrate this functionality into our word-sense disambiguation techniques (both the defining patterns and the semrels) Even further, it seems that matching defining patterns in free text can be used for lexical acquisition Textual material that contains these patterns could conceivably be flagged as providing definitional material which can then be compared to existing definitions to assess whether their use is consistent with these definitions, and if not, at least to flag the inconsistency

The techniques described here can be applied directly to the fields of ontology development and analysis of terminological databases For ontologies, with or without definitions, the methods employed can be used to compare entries in different ontologies based primarily on the relations in the ontology, both hierarchical and other For terminological databases, the methods described here can be used to examine the set of conceptual relations implied by the definitions The definition parsing will facilitate the development of the terminological network in the particular field covered by the database

The componential analysis methods result in a richer semantic network that can be used in other applications. Thus, for example, it is possible to extend the lexical chaining methods described in (Green, 1997), which are based on the semrels used in WordNet. The semrels developed with the componential analysis method would provide additional detail available for application of lexical cohesion methods. In particular, additional relations would permit some structuring within the individual lexical chains, rather than just considering each chain as an amorphous set (Green, 1999).

Finally, we are currently investigating the use of the componential analysis technique for information extraction The technique identifies (from definitions) slots that can be used as slots or fields in template generation Once these slots are identified, we will be attempting to extract slot values from items in large catalog databases (millions of items) In conclusion, it would seem that, instead of a paucity of information allowing us to compare lexical resources, by bringing in the full semantic network of the lexicon, we are overwhelmed with a plethora of data

Acknowledgments

I would like to thank Bonnie Dorr, Christiane Fellbaum, Steve Green, Ed Hovy, Ramesh Krishnamurthy, Bob Krovetz, Thomas Potter, Lucy Vanderwende, and an anonymous reviewer for their comments on an earlier draft of this paper

References

Atkins, B T S (1991) Building a lexicon The contribution of lexicography *International Journal* of Lexicography, 4(3), 167-204

CL Research (1999a) CL Research Demos http://www.clres.com/Demo.html

CL Research (1999b) Dictionary Parsing Project http://www.clres.com/dpp.html

Dolan, W B (1994, 5-9 Aug) Word Sense Ambiguation Clustering Related Senses COLING-94, The 15th International Conference on Computational Linguistics Kyoto, Japan

Green, S J (1997) Automatically generating hypertext by computing semantic similarity [Diss], Toronto, Canada University of Toronto

Green, S J (Sjgreen@mri mq edu au) (1999, 1 June) (Rich semantic networks)

Hovy, E (1998, May) Combining and Standardizing Large-Scale, Practical Ontologies for Machine Translation and Other Uses Language Resources and Evaluation Conference Granada, Spain

Kilgarriff, A (1998) SENSEVAL Home Page http://www.itri.bton.ac.uk/events/senseval/

Krovetz, R (1992, June) Sense-Linking in a Machine Readable Dictionary 30th Annual Meeting of the Association for Computational Linguistics Newark, Delaware Association for Computational Linguistics

Lesk, M (1986) Automatic Sense Disambiguation Using Machine Readable Dictionaries How to Tell a Pine Cone from an Ice Cream Cone Proceedings of SIGDOC

Litkowski, K C (1978) Models of the semantic structure of dictionaries American Journal of Computational Linguistics, Mf 81, 25-74 Litkowski, K C (to appear) SENSEVAL The CL Research Experience Computers and the Humanities

Miller, G A, Beckwith, R, Fellbaum, C, Gross, D, & Miller, K J (1990) Introduction to WordNet An on-line lexical database *International* Journal of Lexicography, 3(4), 235-244

Olsen, M B, Dorr, B J, & Thomas, S C (1998, 28-31 October) Enhancing Automatic Acquisition of Thematic Structure in a Large-Scale Lexicon for Mandarin Chinese Third Conference of the Association for Machine Translation in the Americas, AMTA-98 Langhorne, PA