

Forming an Integrated Lexical Resource for Word Sense Disambiguation

Oi Yee Kwong

Language Information Sciences Research Centre
City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong
rlolivia@cityu.edu.hk

Abstract

This paper reports a full-scale linkage of noun senses between two existing lexical resources, namely WordNet and Roget's Thesaurus, to form an *Integrated Lexical Resource* (ILR) for use in natural language processing (NLP). The linkage was founded on a structurally-based sense-mapping algorithm. About 18,000 nouns with over 30,000 senses were mapped. Although exhaustive verification is impractical, we show that it is reasonable to expect some 70-80% accuracy of the resultant mappings. More importantly, the ILR, which contains enriched lexical information, is readily usable in many NLP tasks. We shall explore some practical use of the ILR in word sense disambiguation (WSD), as WSD notably requires a wide range of lexical information.

1 Introduction

As a repository of lexical information, lexical resources are indispensable for almost every natural language processing (NLP) task. Some NLP tasks, notably word sense disambiguation (WSD), especially require a lot of different types of lexical information to perform well.

There are basically three ways of forming comprehensive resources. One is to start from scratch and encode as much information as desired into a lexical database manually (e.g. Wilks, 1975; Small & Rieger, 1982; Miller et al., 1990). However, this is heavy work, and the time and effort often limit the coverage. Some people have therefore taken an alternative route, by acquiring lexical semantic information semi-automatically from existing resources such as machine-readable dictionaries (e.g. Amsler, 1981; Calzolari, 1988; Chodorow et al., 1985; Vossen et al., 1989) and corpora (e.g. Resnik, 1993; Riloff & Jones, 1999). Nevertheless, this approach may not be entirely satisfactory. A single resource may not contain all types of information. Even if it does (e.g. a comprehensive dictionary might contain thesaural information implicitly in its meaning definitions), the extraction of some information is not always straightforward, rendering the acquired information incomplete or unreliable.

Thus a third and possibly better approach is to combine various existing resources, especially different types of them which are exemplars for representing different types of information, into an integrated repository. By *integration*, we mean that the various resources are linked in some way but the different types of information are preserved in their original structures in individual resources. Such a linkage thus enables us to access the different types of information simultaneously, compatibly, and flexibly.

Although such resource integration has been suggested (e.g. Yarowsky, 1992) or implemented in some way (e.g. Sanfilippo & Poznański, 1992; Knight & Luk, 1994; McHale & Crowter, 1994; Klavans & Tzoukermann, 1995; Chen & Chang, 1998), surprisingly few studies have actually made use of multiple existing lexical resources simultaneously in WSD. Rather, most studies to date which use multiple knowledge sources for WSD are only maximising the exploitation of a single (type of) resource. One study which actually used information distributed across various resources is perhaps

McRoy (1992), but the resources were tailor-made in her study and the linkage between them was manually imposed. We shall explore how an automatically integrated resource could be put into practical use.

In this study, we made use of a simple but effective structurally-based sense-mapping algorithm to link all noun senses shared by two existing lexical resources, namely WordNet and Roget's Thesaurus, to form an *Integrated Lexical Resource* (ILR). The algorithm will be presented in Section 2, followed by a description of the linking process in Section 3. The ILR will be evaluated and its properties discussed in Section 4. Some potential use of the ILR in WSD will be suggested in Section 5, before we conclude in Section 6.

2 A Structurally-Based Sense-Mapping Algorithm

An obvious interface between different resources is the senses in common to them. Hence combining multiple resources can be treated as a task of sense mapping among the resources. However, this may not be straightforward. On the one hand, resources notoriously differ in sense distinction and granularity. On the other hand, the type(s) of information associated with the senses in different resources can be very different (but that is what we want to preserve) which makes the linkage more tricky. For instance, McHale and Crowter (1994) and McHale (1995) tried to map senses in the Longman Dictionary of Contemporary English (LDOCE) to the 1,024 categories in the Roget's International Thesaurus. Only moderate accuracy (about 63%) was obtained. Kwong (1998a) suggested that the knowledge gap between a conventional dictionary and a thesaurus might be too big to be crossed in one step. The effectiveness of sense mapping depends on how well one can work out the semantic relationships (implicitly embedded) among the words in a dictionary definition with respect to the thesaural classes. This is far too constrained by the actual dictionary definitions. For instance, McHale and Crowter (1994) also found that results varied considerably with different dictionaries. Hence, on the one hand, the mapping might be better done with an intermediate step. Such a step could be built on work like Sanfilippo and Poznański (1992), Knight and Luk (1994), and Chen and Chang (1998). On the other hand, a more robust algorithm should depend only minimally on the lengths and the exact wordings of the dictionary definitions.

Kwong (1998a; 1998b) thus designed a structurally-based sense-mapping algorithm, which exploits the structural inter-relatedness among resources, and only minimally depends on specific contents of individual resources. The algorithm maps senses from a dictionary to a thesaurus via a *mediator*. A mediator is another kind of lexical resource which shares structural properties with both a dictionary and a thesaurus, an exemplar of which is WordNet (WN). Meanwhile, LDOCE and Roget's Thesaurus (ROGET) were used as exemplars for dictionary and thesaurus respectively. Thus words are organised in alphabetical order in LDOCE, as in other conventional dictionaries. The senses are listed after each entry, in the form of text definitions. WN groups words into sets of synonyms ("synsets"), with an optional textual gloss. These synsets form the nodes of a taxonomic hierarchy. In ROGET, each semantic head comes with a number, under which words are first assorted by part of speech and then grouped into paragraphs according to broad conceptual categories. On the other hand, heads are in turn grouped into sections and then into classes, resulting in a hierarchical organisation. Hence we can exploit the structural inter-relatedness to map senses from the dictionary to the mediator and from the mediator to the thesaurus separately. With the mediator as the backbone and assuming transitivity, a *mapping chain* from the dictionary to the thesaurus through the mediator can then be derived.

A schema for the structural inter-relatedness among the various types of resource, illustrated with LDOCE (dictionary), WN (mediator) and ROGET (thesaurus), is shown in Figure 1. The arrows in the figure indicate the direction of mapping. In the middle of Figure 1 is the structure of WN, a hierarchy with synsets at the nodes. A word x_2 has a sense as in synset X . On the one hand, since classes in ROGET are more general, we expect to find not only words in synset X , but also those in the coordinate synsets (i.e. M and P , with words m_1, m_2, p_1, p_2 , etc.) and the superordinate synsets (i.e. C and A , with words c_1, c_2 , etc.) in the same ROGET paragraph. In other words, the thesaurus

class to which x_2 belongs should overlap with $X \cup M \cup P \cup C \cup A$. On the other hand, the LDOCE definition corresponding to the sense of synset X (i.e. $\sim D_x$) is expected to be similar to the textual gloss of synset X (i.e. $\sim Gl(X)$). Also, since it is not unusual for dictionary definitions to be phrased with synonyms or superordinate terms, we would also expect to find words from X and C , or even A , in the LDOCE definition, i.e., $D_x \approx Gl(X)$ and $D_x \cap (X \cup C \cup A) \neq \emptyset$.

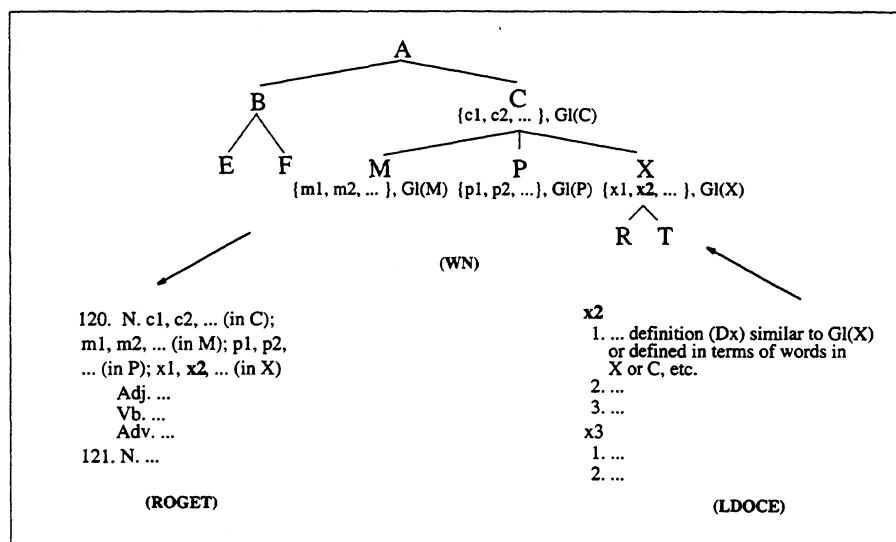


Figure 1 Schema of Inter-relatedness among Resources

The algorithm, which only handles noun senses at present, is summarised below. Again, we use ROGET, LDOCE and WN as exemplars for the various resource types.

Step 1: From LDOCE, get the sense definitions $D_1 \dots D_i$ under the word W .

Step 2: From WN, get all the synsets $S_n\{w_{n1}, w_{n2}, \dots\}$ such that $W \in S_n$. Collect the corresponding gloss definitions $Gl(S_n)$, the chain of hypernym synsets $Hyp(S_n)$, and the coordinate synsets $Co(S_n)$, if available.

Step 3: Convert the above data to flattened lists of words, removing punctuations, function words, and repeated words from the definitions (LDOCE definitions and WN glosses). Compute a similarity score matrix A for the LDOCE senses and the WN synsets, where $A(i,j)$ is the similarity score between the i^{th} LDOCE sense (i.e. D_i) and the j^{th} WN synset (i.e. S_j). It is a weighted sum of the overlaps between the LDOCE definition and the WN synset, hypernym chain, and gloss respectively:

$$A(i,j) = a_1 |D_i \cap S_j| + a_2 |D_i \cap Hyp(S_j)| + a_3 |D_i \cap Gl(S_j)|$$

The mappings from LDOCE to WN can be obtained by choosing the maximum element in each row in matrix A .

Step 4: From ROGET, get all the paragraphs $P_m\{w_{m1}, w_{m2}, \dots\}$ such that $W \in P_m$.

Step 5: As in Step 3, compute a similarity score matrix B for the WN synsets and the ROGET paragraphs, where $B(j,k)$ is the similarity score between the j^{th} WN synset (i.e. S_j) and the k^{th} ROGET paragraph (i.e. P_k), and is a weighted sum of the overlaps between the ROGET paragraph and the WN synset, hypernym chain, and coordinate terms respectively:

$$B(j,k) = b_1 |S_j \cap P_k| + b_2 |Hyp(S_j) \cap P_k| + b_3 |Co(S_j) \cap P_k|$$

The mappings from WN to ROGET can be obtained by choosing the maximum element in each row in matrix B .

Step 6: For $i = 1$ to t (i.e. each LDOCE sense), find $\max(A(i,j))$ from matrix A . Then trace from matrix B the j^{th} row and find $\max(B(j,k))$. The final mapping is then established between D_i and P_k (or more precisely, the head of P_k).

3 Forming an Integrated Lexical Resource

In this work, we capitalise on the success of the direct mapping between the mediator and the thesaurus with the above algorithm, and do a full-scale mapping of noun senses between WordNet and Roget's Thesaurus to start up an Integrated Lexical Resource (ILR).

WordNet 1.6 (WN16) was used. It was released by Princeton in March 1998. The Prolog version of WN16 has a total of 94,474 distinct nouns¹, which amount to 116,364 senses altogether.

For Roget's Thesaurus, we used a joint version of the 1911 Crowell Roget's Thesaurus (ROG11) and the 1987 Penguin Roget's Thesaurus (ROG87)², collectively called ROGET. The ROG87 index itself contains 20,024 distinct nouns and 47,416 noun senses. The joint version, ROGET, has 35,718 distinct nouns and 75,013 senses.

The linkage between WN16 and ROGET was done with steps 2, 4, and 5 of the algorithm described in Section 2. However, some adjustments were made to the algorithm. First, the hypernym chain from WN16 (i.e. $Hyp(S_j)$ in step 2) was limited to two levels from the target synset up the hierarchy (instead of a complete chain to the top node). This was because we observed that high level nodes were often too general to be helpful. Second, heads (with all paragraphs in them) instead of individual paragraphs in ROGET were taken as basic units, since we had actually disturbed the paragraph structure when we supplemented ROG87 with ROG11. In addition, the resultant mappings must have scores greater than 1, which means there must be at least one other word in common apart from the target word itself, otherwise the link would be left void. Ties in scores were preserved, allowing multiple links from the source sense to the target senses.

4 Evaluation

In this section, we will evaluate the Integrated Lexical Resource from a randomly selected sample of nouns. We will probe the coverage of the ILR in Section 4.1 by seeing how compatible the two component resources are. Then we will explain the performance measures in Section 4.2, and report the results and observations in Section 4.3.

4.1 Compatibility of WN16 and ROGET

The full-scale mapping of nouns between WN16 and ROGET suggests the compatibility of the two resources as shown in Figure 2. The shaded area (including the netted area) represents the senses in

¹ In making the estimates, capital letters are disregarded. So "March", "march", and "MArch" will be counted as different senses of the same word, just as they are displayed from the online interface.

² Although ROG11 is machine-readable, ROG87 nevertheless has a bigger and more up-to-date collection of words. So we did some pre-processing to join the two versions as in Kwong (2000).

WN16 of words which appear in both resources, and the netted area represents the senses among them which were mapped.

We can see that while the two resources differ considerably in collection size, about 90% of the nouns common to both resources (i.e. about 18,000 nouns with over 30,000 senses) could be linked in some way. The difference in size, however, was observed to arise from the peculiar contents of individual resources. For instance, WN16 contains a lot of specialised words (e.g. “genus Amphioxus”, “genus Salpa”, “order Coniferales”, “order Taxales”, etc.) which are absent from ROGET. On the other hand, ROGET contains many multiple-word nominals (e.g. “amateur dramatics”, “ancient monument”, etc.) which are not found in WN16. In other words, although the number of linked words is just about half of the ROGET collection and only about one-fifth of WN16, these shared words are nevertheless among the commonly used words, unlike the above peculiar examples. Therefore we believe that the results from the linkage would form a usefully enriched resource for WSD, despite its limited size.

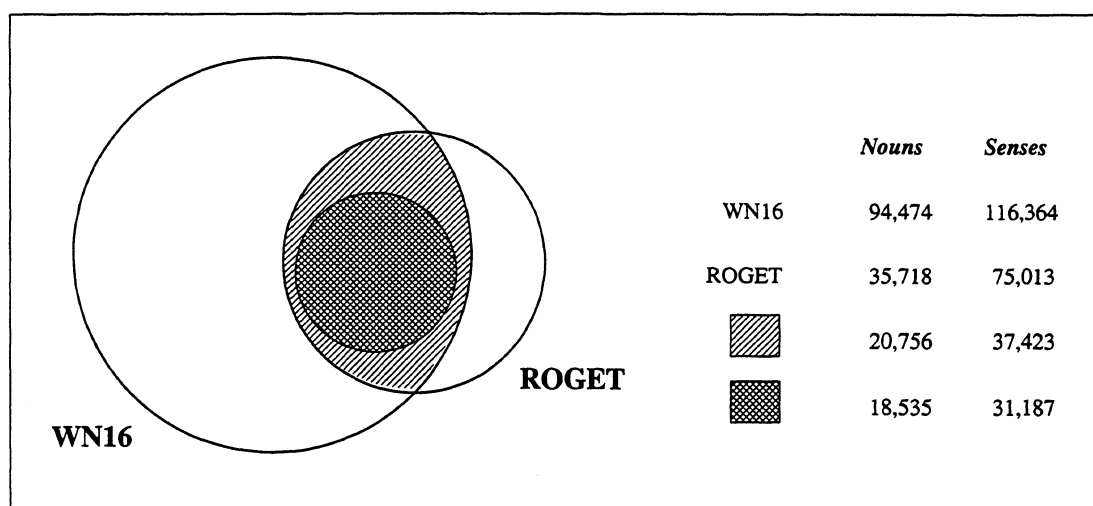


Figure 2 *Compatibility of WN16 and ROGET*

4.2 Performance Measures

In this section, we introduce the measures we use to evaluate the performance of the sense-mapping algorithm described above. Although in the current study we have focussed on the *direct mapping* from the mediator to the thesaurus, the evaluation criteria described below also apply to the direct mapping from the dictionary to the mediator, and most of them to the *indirect mapping* from the dictionary to the thesaurus.

It is not at all clear what an objective measure of accuracy for sense mapping between resources should be, mainly because resources do not necessarily have a one-to-one correspondence between the senses they contain. For this reason, we tend to be lenient in judging the appropriateness of the mapping. We therefore allow various degrees of success, e.g.:

- *Accurate*: mapping can be judged as correct right away.
- *Alternatively Accurate*: mapping is not related directly by the genus term, BUT can be related by other properties conveyed by the sense.
- *Barely Accurate*: mapping does not fit the above two criteria, but the overlapping words are convincing enough AND there is apparently no better match possible.

However, though we have these rules of thumb for our judgement, with borderline cases we have but our intuition to count on; and since it is quite difficult to distinguish such varying degrees of accuracy in practice, we will group them into a single category called *Accurately Mapped (A)*.

On the contrary, errors can be more easily marked as different types. For the direct mapping, we will identify four types of error:

- *Incorrectly Mapped (I)*: mapping is unsatisfactory AND a better alternative can be found.
- *Forced Error (F)*: mapping is unsatisfactory AND no better alternative is available.
- *Unmapped-a (Ua)*: mapping is empty BUT a mapping is expected.
- *Unmapped-b (Ub)*: mapping is empty AND none of the alternatives is good enough.

The error types for the direct mapping are summarised in Table 1. We can see that *I* and *Ua* are *misses* because they fail to identify the expected target. On the other hand, *F* and *Ub* are *false alarms* because they produce something when nothing is expected.

Target Exists	Mapping Outcome	
	Wrong Match	No Match
Yes	<i>Incorrectly Mapped (I)</i>	<i>Unmapped-a (Ua)</i>
No	<i>Forced Error (F)</i>	<i>Unmapped-b (Ub)</i>

Table 1 *Error Types in Direct Sense Mapping*

In addition, we will also present the results using more general numeric measures such as *precision*, *recall*, and *coverage*. They are defined for the direct mappings as follows:

$$precision = \frac{A}{A + I + F}$$

$$recall = \frac{A}{A + I + F + Ua + Ub}$$

$$coverage = \frac{A + I + F}{A + I + F + Ua + Ub}$$

4.3 Sample Inspection and Further Observations

Although ideally we should make sure the mappings are all correct before putting the integrated resource to use, unfortunately full-scale checking is impractical given the huge amount of data. However, we could exploit the test results for the sense-mapping algorithm as reported in Kwong (1998a; 1998b) to estimate the success of the full-scale mapping. Since the samples in these tests are among the most common words, we believe that they are representative of most words in the rest of

the lexicon as far as the mapping effectiveness is concerned. Referring to previous results, it is thus reasonable to expect some 75% success rate for the full-scale linkage.

To investigate this claim further, we inspected the mappings for a random sample of nouns. These 19 nouns, listed in Table 2, were taken from a randomly selected excerpt from the Brown Corpus³. Columns W and R in Table 2 indicate the number of senses each noun has in WN16 and ROGET respectively.

Noun	W	R	Noun	W	R	Noun	W	R	Noun	W	R
concert	1	6	point	24	21	rhythm	5	10	man	11	8
score	11	14	joust	1	1	wit	3	8	baton	2	2
fate	3	7	tune	1	2	kind	1	2	technique	2	4
time	10	9	misfortune	2	4	performance	5	11	style	8	18
thrust	5	7	texture	4	4	orchestra	2	4			

Table 2 *Nouns for Inspection of Full-Scale Mapping*

Table 3 shows the performance of mapping the above samples according to the measures described in Section 4.2. The results corroborate most of the findings from earlier pilot experiments of the algorithm as in Kwong (1998a; 1998b). For instance, nouns with five or fewer WN16 senses are most prevalent among the samples. The mapping accuracy is about 70% on average, which is close to our expectation; and mapping performance deteriorates for highly polysemous words (i.e. words with 11 or more WN senses, in our definition). Meanwhile, the errors, upon further analysis, illustrate again that the mapping was mostly let down by the difference in content between resources. In terms of coverage, about 90% of the senses were mapped in some way, and the precision is over 75%. Putting it in perspective, as only a very small portion of the words in WN are highly polysemous, the overall precision for the mappings in the ILR should be even better.

Performance Measure	WN16 → ROGET
<i>Accurately Mapped (A)</i>	68.32%
<i>Incorrectly Mapped (I)</i>	6.93%
<i>Unmapped-a (Ua)</i>	1.98%
<i>Unmapped-b (Ub)</i>	7.92%
<i>Forced Error (F)</i>	14.85%
<i>Precision</i>	75.82%
<i>Recall</i>	68.32%
<i>Coverage</i>	90.10%

Table 3 *Results for mapping the samples in Table 2 from WN16 to ROGET*

In addition, the current analysis uncovers further interesting differences between WN16 and ROGET. For instance, the construction of WN16 was to a certain extent corpus-based, and it contains word senses which were derived from actual usages of words in the Brown Corpus. As a consequence, there may be some non-central (and perhaps mildly technical) senses in WN16 which

³ *Brown Corpus c01*, 4th paragraph.

one might not find in an ordinary resource, e.g. the "fourth dimension" sense of "time" is found in WN16, but not ROGET.

Apart from the origin of senses, WN16 and ROGET also seem to differ in the aspects of senses they emphasise. The latter difference may account for the weak correlation of the latter with other resources in terms of the number of senses for words (Kwong, 1998b). Such a disproportion of senses across resources is also obvious in the current analysis. For example, "concert" has only one sense in WN16, namely "a performance of music". But it is under six heads in ROGET: 24(Agreement), 181(Concurrence), 410(Melody), 412(Music), 414(Musical instruments), and 706(Cooperation) -- everything but "performance"! This variation of focus thus leads to unexpected mappings and demands some flexibility in our judgement of appropriateness. For example, given that "concert" is not under "performance" in ROGET, we might need to evaluate the mapping with respect to other properties of the sense. In this case, the next relevant property is apparently related to "music", and we will be at the crossroads again since there are three alternatives (heads 410, 412, and 414). However, in practice we will treat all of them as acceptable targets.

5 Using the Integrated Lexical Resource

As mentioned earlier, few studies have actually made use of multiple existing lexical resources simultaneously in WSD. Without real application of the integrated resource, it defeats the purpose of combining resources. Here we suggest some practical use of our ILR in WSD.

It happens that both component resources in our ILR (WN16 and ROGET) have a hierarchical classification of senses, although the classificatory criteria are quite different. With the ILR, the two classifications can be used together compatibly. We can thus take two senses as defined in WN16 but compare them with respect to the ROGET classification, as in the example below.

Figure 3 shows the mapping (indicated by bullet points) of "performance" from WN16 to ROGET. The leftmost column shows the WN16 senses while the top row shows the ROGET heads of "performance". On the one hand, we see that the last three WN16 senses of "performance" are closely related meanings in the ROGET classification by being under the same ROGET head. However, these senses are distinguished in the WN16 classification and in fact, they relate to head 676 in ROGET via different groups of words. On the other hand, while sense 2 to 5 in WN16 are similar in the sense that they all descend from the top node of "human action", sense 2 goes to a different ROGET head from the others. Hence senses which are close in one classificatory system may be distant in another, depending on what semantic relations we are considering. More importantly, WSD not only requires these different types of semantic relation as disambiguation evidence, but also requires them to be applied appropriately.

	157 (Effect)	164 (Production)	413 (Musician)	551 (Representation)	594 (Drama)	676 (Action)	725 (Completion)	768 (Observance)	876 (Celebration)	917 (Duty)	988 (Ritual)
1. dramatic / musical entertainment					•						
2. act of presenting entertainment				•	•				•		
3. act of doing something						•					
4. process of functioning						•					
5. recognized accomplishment						•					

Figure 3 Mapping Results for "performance" in WN16 → ROGET

We can adopt a quantitative way of doing such cross-resource similarity comparison. For instance, we first define $M(s)$ as "the set of ROGET heads that sense s is mapped to", and $S(h)$ as "the ROGET section (one level above heads) subsuming head h ". We can then define a function for measuring the similarity between WN16 senses s_1 and s_2 as in ROGET, like this:

$$Sim(s_1, s_2) = \begin{cases} 2 & \text{if } M(s_1) \cap M(s_2) \neq \emptyset \\ 1 & \text{if } M(s_1) \cap M(s_2) = \emptyset \\ & \text{and } h_1 \in M(s_1), h_2 \in M(s_2), S(h_1) = S(h_2) \\ 0 & \text{otherwise} \end{cases}$$

Since the ROGET hierarchy is a relatively flat one, a step function should be appropriate. The actual values of the function could be determined empirically, as long as they correlate with the similarity between two senses with respect to the ROGET hierarchy. Applying the above function to "performance", we get:

$$\begin{aligned} Sim(\text{performance}/1, \text{performance}/2) &= 2 \\ Sim(\text{performance}/1, \text{performance}/4) &= 0 \end{aligned}$$

The ability to quantify such comparisons means that we can combine different resources into a single sense inventory, but preserve the structures of individual resources to allow the flexible use of different types of information for WSD. This is the most significant property of the ILR, and the ultimate motivation for integrating multiple existing resources.

In actual WSD, the above comparison is normally between senses of different words. However, comparing senses of the same word, our example serves to illustrate another potential use of the mapping results, in terms of WSD evaluation. Resnik and Yarowsky (1997) suggested that wrong answers from WSD systems should be given different penalties depending on the closeness of the wrong answer to the expected one. Our sense-mapping matrix can therefore serve as a kind of confusion matrix, and the similarity function above offers an alternative way for determining the closeness of senses. In the case of "performance", a system which fails to distinguish between senses 1 and 2 should therefore receive less penalty than one which fails between senses 1 and 4.

6 Conclusion

We have performed a large-scale automatic sense linkage between WordNet 1.6 and Roget's Thesaurus to form an integrated resource, using a structurally-based sense-mapping algorithm. About 18,000 nouns with over 30,000 senses were mapped. The mapping accuracy is about 70%, while failure is often due to senses present in one resource but absent in another. The accuracy figure thus suggests how reliable the mappings will be for subsequent use. The integrated resource is readily usable, and we have discussed some potential use of it in WSD. Future work includes the full use of the mapping algorithm to incorporate resources of the dictionary type into the ILR, and the application of the ILR in WSD studies.

Acknowledgements

This work was done at the Computer Laboratory, University of Cambridge. The author would like to thank Prof. Karen Sparck Jones for her advice and comments. The work was financially supported by the Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom, the Cambridge Commonwealth Trust, Downing College, and the Croucher Foundation.

References

- Calzolari, N. 1988. The dictionary and the thesaurus can be combined. In M.W. Evens (ed.), *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*. Cambridge University Press.
- Chen, J.N. and Chang, J.S. 1998. Topical clustering of MRD senses based on information retrieval techniques. *Computational Linguistics*, 24(1): 61-95.
- Chodorow, M.S., Byrd, R.J., and Heidorn, G.E. 1985. Extracting semantic hierarchies from a large on-line dictionary. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics (ACL '85)*, pages 299-304, Chicago.
- Klavans, J. and Tzoukermann, E. 1995. Combining corpus and machine-readable dictionary data for building bilingual lexicons. *Machine Translation*, 10: 185-218.
- Knight, K. and Luk, S.K. 1994. Building a large-scale knowledge base for machine translation. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, pages 773-778, Seattle, Washington.
- Kwong, O.Y. 1998a. Bridging the Gap between Dictionary and Thesaurus. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL '98)*, pages 1487-1489, Montréal, Canada.
- Kwong, O.Y. 1998b. Aligning WordNet with Additional Lexical Resources. In *Proceedings of the Workshop on Usage of WordNet in Natural Language Processing Systems*, Montréal, Canada.
- Kwong, O.Y. 2000. *Word Sense Selection in Texts: An Integrated Model*. Technical Report No.504, Computer Laboratory, University of Cambridge, U.K.
- McHale, M.L. 1995. *Combining Machine-Readable Lexical Resources with a Principle-Based Parser*, chapter 6. Ph.D. Dissertation, Syracuse University, New York.
- McHale, M.L. and Crowter, J.J. 1994. *Constructing a lexicon from a machine readable dictionary*. Technical report RL-TR-94-178, Rome Laboratory, Griffiss Air Force Base, New York.
- McRoy, S. 1992. Using multiple knowledge sources for word sense disambiguation. *Computational Linguistics*, 18(1): 1-30.
- Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K.J. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4): 235-244.
- Resnik, P. 1993. *Selection and Information: A Class-Based Approach to Lexical Relationships*. Doctoral Dissertation, Department of Computer and Information Science, University of Pennsylvania.
- Resnik, P. and Yarowsky, D. 1997. A perspective on word sense disambiguation methods and their evaluation. In *Proceedings of SIGLEX '97 Workshop: Tagging Text with Lexical Semantics: Why, What, and How?*, pages 79-86, Washington, D.C.
- Riloff, E. and Jones, R. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pages 474-479, Orlando, Florida.
- Sanfilippo, A. and Poznański, V. 1992. The acquisition of lexical knowledge from combined machine-readable dictionary sources. In *Proceedings of the 3rd Conference on Applied Natural Language Processing*, pages 80-87, Trento, Italy.
- Small, S. and Rieger, C. 1982. Parsing and comprehending with word experts (A theory and its realization). In W.G. Lehnert and M.H. Ringle (eds.), *Strategies for Natural Language Processing*. Lawrence Erlbaum Associates, New Jersey.

- Vossen, P., Meijs, W., and den Broeder, M. 1989. Meaning and structure in dictionary definitions. In B. Boguraev and T. Briscoe (eds.), *Computational Lexicography for Natural Language Processing*. Longman Group, Essex, England.
- Wilks, Y. 1975. A preferential, pattern-seeking, semantics for natural language inference. *Artificial Intelligence*, 6: 53-74.
- Yarowsky, D. 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, pages 454-460, Nantes, France.

