Mapping Dictionaries: A Spreading Activation Approach

Nancy M. IDE[1] and **Jean VERONIS**[1,2]

[1] Department of Computer Science
VASSAR COLLEGE
Poughkeepsie, New York 12601 (U.S.A.)

[2] Groupe Représentation et Traitement des Connaissances
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
31, Ch. Joseph Aiguier
13402 Marseille Cedex 09 (France)

New OED Conference, Waterloo, October 1990

Abstract

In this paper we apply a spreading activation algorithm to the problem of mapping senses between machine readable dictionaries, which is required in order to combined information extracted from them. The algorithm is run over networks automatically constructed from dictionary definition texts. On a sample corpus of sense definitions, our strategy correctly identifies the corresponding homograph in a second dictionary in 97% of the cases, and the correct sense within that homograph in 90% of the cases. This result is a substantial improvement over previously proposed strategies, such as Lesk's method. It also demonstrates that spreading activation strategies, which have rarely been used in computational lexicography, are one means to exploit the full potential of relations implicitly encoded in machine-readable dictionaries.

The authors would like to thank Lisa Lassek for her contribution to the work of this project.

1. Introduction

It is widely recognized that machine readable dictionaries are a rich source of semantic information about word senses, and several researchers are working on automatic procedures to identify and extract this information for use in creating lexical knowledge bases (see, for instance, Amsler, 1980; Calzolari, 1984; Chodorow *et al.,* 1985; Markowitz *et al.,* 1986; Byrd *et al.,* 1987; Slator and Wilks, 1987; Véronis and Ide, 1990). However, dictionaries differ considerably in the amount and kinds of information they include, and no single dictionary seems to be complete. It is therefore advantageous to combine information extracted from several dictionaries in order to create a comprehensive knowledge base. For example, it has been shown that combining taxonomic information automatically extracted from several dictionaries significantly improves the completeness and precision of the resulting taxonomic tree (Ide and Véronis, 1990a; Véronis *et al.,* 1990).

When information from multiple dictionaries is to be combined, we encounter what has been called the "mapping problem": that is, the senses given for a word in one dictionary must be mapped onto those given in each of the others, in order to determine which information applies to a common concept. Obviously, automated means to accomplish bilateral dictionary mappings is desirable, and at least two automated sense mapping strategies have been proposed (Chodorow, Ravin, and Sachar, 1988; Byrd, 1989).

In this paper, we describe a means to automatically map senses between machine-readable dictionaries by applying a spreading activation algorithm to networks created from one dictionary's definition texts. We have already seen promising results from the application of the spreading activation approach to word sense disambiguation (Ide and Véronis, 1990b; Véronis and Ide, 1990). The success of the strategy relies upon common links within the network for the words used to "prime" the network when the spreading activation algorithm is applied. For word sense disambiguation, the input consists of a word to be disambiguated and one or more other words with which it appears in context; for sense mapping, input consists of words in a sense definition from a dictionary other than the one from which the network was created. Because we can expect a greater degree of connectivity among words within a sense definition text within the network than for words in context, our strategy should be even more successful for sense mapping than for word sense disambiguation.

The work described in this paper has been carried out in the context of a joint project of Vassar College and the Groupe Représentation et Traitement des Connaissances of the Centre National de la Recherche Scientifique (CNRS), which is concerned with the construction and exploitation of a large lexical data base of English and French. At present, the Vassar/CNRS data base includes, through the courtesy of several editors and research institutions, several English and French dictionaries (the *Collins English Dictionary,* the *Oxford Advanced Learner's Dictionary,* the *COBUILD Dictionary,* the *Longman's Dictionary of Contemporary English*, the*Webster's 9th Dictionary,* and the *ZYZOMYS* CD-ROM dictionary from Hachette Publishers) as well as several other lexical and textual materials (the *Brown Corpus of American English,* the CNRS *BDLex* data base, the *MRC Psycholinguistic Data Base,* etc.). A broad description of the database and the activities within the Vassar/CNRS project appears in Véronis *et al.* (1990).

2. The mapping problem

Sense distinctions in published dictionaries differ widely in their number and form, and thus the problem of mapping dictionaries is not straightforward. For instance, monolingual dictionaries for use by native speakers often specify subtle distinctions among senses and include rare or obsolete senses. Learner's dictionaries, on the other hand, typically provide fewer, more clearly-cut sense distinctions. These differences are clear in the definitions of *ash* from the *Collins English Dictionary (CED),* which is intended for native speakers, and the *Oxford Advanced Learner's Dictionary (OALD)* given in figure 1.

CED:

ash 1
1. the nonvolatile products and residue formed when matter is burnt.
2. any of certain compounds formed by burning.
3. fine particles of lava thrown out by an erupting volcano.
4. a light silvery grey colour, often with a brownish tinge.
ash 2
1. any oleaceous tree of the genus *Fraxinus*, esp. *F. excelsior* of Europe and Asia, having compound leaves, clusters of small greenish flowers, and winged seeds.
2. the close-grained durable wood of any of these trees, used for tool handles, etc.
3. any of several trees resembling the ash, such as the mountain ash
ash 3
the digraph æ, as in Old English, representing a front vowel approximately like that of the *a* in Modern English *hat*. The character is also used to represent this sound in the International Phonetic Alphabet.

OALD:

ash 1
forest tree with silver-grey bark and hard, tough wood; wood of this tree
ash 2
1. powder that remains after [something] has burnt
2. the burnt (=cremated) remains of a human body

Figure 1. Definition of "ash" from the CED and OALD

These two definitions demonstrate many of the difficulties inherent in mapping dictionaries. In the best case, a sense mapping between two dictionaries for a given sense is *one-to-one,* in which case there is a direct mapping between the sense in one dictionary and some sense of the same word in the other. This is the case for *ash 1.1* in the *CED* and *ash 2.1* in the *OALD.* In other cases, the mapping may be *one-to-many,* where a single sense in one dictionary maps onto two or more in the target dictionary, or the reverse, where the mapping is *many-to-one*. If we do not take the semicolon in the *OALD's ash 1* as dividing this sense into two senses, this sense provides an example of a one-to-many mapping because it maps onto both *ash 2.1* and *ash 2.2* in the *CED.* (Note that in our experiments, we do take the semicolon as dividing senses.) Finally, there can be a *one-to-zero* mapping, where there is no corresponding sense in the target dictionary, which is the case for *ash 2.2* from the *OALD* and, mapping from the *CED* to the *OALD,* for *CED* senses *ash 1.3, ash 1.4, ash 2.3,* and *ash 3.* Occasionally, a mapping may be even more difficult, where a given sense corresponds only partially to a single sense or perhaps corresponds partially to two or three senses in the target dictionary, without directly overlapping any one of them.

Previously proposed solutions to the mapping problem derive from Lesk's (1986) sense disambiguation method, which computes the degree of overlap--that is, number of shared words--in definition texts. Applied straightforwardly to the mapping problem, this procedure determines the senses of a word from each of two dictionaries with the greatest number of overlaps. Although this strategy has been applied with some success to the mapping of synonym dictionaries (Chodorow, Ravin, and Sachar, 1988), it appears that its success is limited for mapping ordinary dictionaries, with fewer than 50% correct mappings (Byrd, 1989). Byrd proposed a similar strategy which computes overlaps between *dictionary sense property vectors*, which for each sense of a given word include the words in its definition as well as other information about these words extracted from the definition text, such as typical subject, object, and instrument; subject and object selectional restrictions; synonyms; etc. The degree of overlap is computed by matching words property-by-property, and thus the mapping criteria are more constrained than in the earlier approach. To date, no report of the results of applying this method has been published.

All of these suggested mapping procedures, even those using extracted information, rely for their success on the presence of shared words in corresponding senses. However, these methods typically succeed due to the presence of only one or two shared words, and thus the relation between senses is often very tenuous. Another problem arises when the same number of overlaps exists with more than one sense, in which case it is not possible to determine which is the correct mapping. These methods also fail to take into account less immediate relationships between words.

As a result, they will not determine, for instance, the sense of *cape* in the *CED* which corresponds to *cape2* ("a high point of land going out to the sea") in the *OALD,* because the two do not share any words in common. However, a strategy which takes into account a longer path through definitions will find that the word *headland,* which appears in the corresponding definition in the *CED,* contains the words *land* and *sea* in its definition, and both these words appear in the *OALD* definition. Longer paths may also reveal additional links between senses, which can help refine information about overlaps.

3. Network topology

We demonstrate an automatic method for sense mapping that does not demand the presence of shared words in the definition texts of corresponding senses, but instead utilizes a complex network of words and senses automatically constructed from dictionary definition texts. In our experiments so far, definition texts from the *CED* have been used to build the network.

The definition network is built by a simple, straightforward, automatic procedure that does not require hand coding or sophisticated analysis to extract information from definitions (which may not be present even if such analysis is applied). The network is constructed by first extracting the definition text for each word in the *CED.* The definition texts are pre-processed to remove function and other common or problematic words (for instance, words with 70 or 80 senses of their own) and all remaining words are morphologically normalized (figure 2). The network is constructed with a simple program that scans the pre-processed definition texts and creates links among words and senses. Within the network, each lexical entry is represented by a complex grouping of nodes consisting of a central node (or *word node*) which represents the lexical entry itself. This node is connected to a number of *sense nodes* that represent the different senses (definitions) of this word in the *CED.* Thus for the lexical entry *ash,* eight sense nodes would exist, one corresponding to each of the eight senses of *ash* given in the *CED.* Each of these sense nodes is connected to the word nodes representing the words that appear in its definition. These words are, in turn, connected to sense nodes according to their definitions within the *CED*, etc. In addition, for each connection from a node *i* to a node *j,* there is a reciprocal connection from node *j* to node *i.* A small portion of the network is pictured in figure 3.

ash
1.1 nonvolatile product residue matter burn
1.2 compound burn
1.3 particle lava throw erupt volcano
1.4 light silvery grey colour brown tinge
2.1 oleaceous tree compound leave cluster green flower wing seed
2.2 close grain durable wood tree tool handle
2.3 tree resemble ash mountain ash
3.1 digraph english front vowel modern english hat character sound international phonetic alphabet

Figure 2. Pre-processed definition for the word "ash"

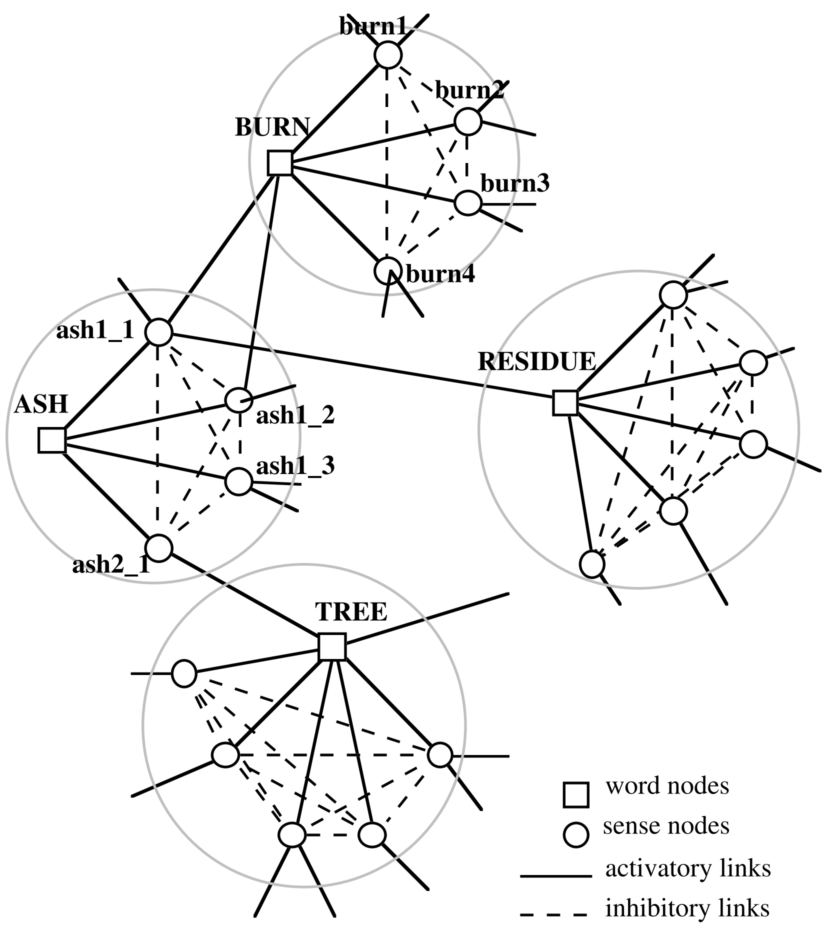

Figure 3. Topology of the network

We assume that there is a semantic relation between a word and the words used to define it. Therefore, we can view the structure we build from definition texts in the *CED* as a highly complex, interconnected network in which semantically related words are connected by one or more paths within the network. The more related the words, the more connected they should be, both by a greater number of connections and by shorter pathlengths within those connections. We rely on these assumptions to uncover both immediate and less than immediate semantic relations among words and senses: our method does not depend, as Lesk-based strategies do, on a direct connection between two words (reflecting an overlap in the wording of two definition texts), and thus more remote relations can be detected. Thus, even when no overlapping words exist in corresponding definition texts, a relationship between the definition words may be discoverable.

Ideally, the network we build would include the entire dictionary. However, the resulting network for the *CED,* which contains 90,000 words, would be enormous, and so for practical reasons we currently limit the size of the network to a few thousand nodes and 10 to 40 thousand transitions. We do this by building only the portion of the network that represents the input words, the words in their definitions, and the words in these words' definitions in turn (as well as the intervening sense nodes). Thus for each set of input words, a potentially different network is constructed. Obviously, increasing the network's coverage of the dictionary would substantially increase its interconnectivity, which could improve our results but which may also obscure straightforward semantic relations by being "overconnected." We are currently working on a large-scale implementation which would enable us to experiment with the use of a network built from the entire dictionary, or larger sub-portions of it.

4. The spreading activation strategy

The spreading activation strategy requires that links among nodes in the network be either *activatory* or *inhibitory*. Practically speaking, activatory links are those with a positive weight (in our model, a number between 0 and +1) and inhibitory links are those with a negative weight (0 to -1). The effect of these weights when the spreading activation algorithm is applied is described below. In our model, the connections from word to sense nodes and from sense to word nodes are activatory, and the different sense nodes for a given word are interconnected by inhibitory links, as shown in figure 3.

Again, we employ a simple program to apply the spreading activation algorithm over the network. The activation model we use is derived from McClelland and Rumelhart (1981). In general terms, the procedure works by first activating the nodes within the network corresponding to the input words (in our case, a headword and words from one of its *OALD* senses; see section 5.1). Within the network at any time, each node has an *activation level*, initially set to 0 for all nodes in the network. The activation initially applied to the input nodes gives these nodes an activation level of 1 in our implementation. Then, each input word node sends activation to its sense nodes, which in turn send activation to the word nodes to which they are connected, and so on throughout the network for a number of cycles. The amount of activation sent by a node to its neighbor is a product of the activation level of that node and the weight on the link between the node and its neighbor. Sense nodes for the same word are interconnected by inhibitory links, which have negative weights, and thus send negative activation or *inhibition* to one another. These nodes can be considered to be "in competition", since the more active one becomes, the more it will tend to decrease its neighbors' activation levels. Word and sense nodes also receive *feedback* via the reciprocal links from the nodes to which they are connected. The total amount of activation received by any node will be the sum of the activation and inhibition received from all of its neighbors. Because the dictionary network can be heavily interconnected, a given node may receive input from a large number of sources.

The spreading activation algorithm proceeds by repeating this process, in which activation is applied to input words and activation is subsequently sent from node to node throughout the network, over a number of cycles. As the cycles progress, feedback and inhibition cooperate in a "winner-take-all" strategy to activate increasingly related word and sense nodes and deactivate the unrelated or weakly related nodes. Eventually, after a few dozen cycles, the network stabilizes in a configuration where only the sense nodes with the strongest relations to other nodes in the network are activated. Because of the "winner-take-all" strategy, at most one sense node per word will ultimately be activated.

We can describe the activation model in mathematical terms as follows. At any instant t, a node i in the network has an activation level $a_i(t)$ with a real value. If a node has a positive activation level, it is said to be *active*. If the node receives no input from its neighbors, it tends to return to a *resting level* r_i, more or less quickly depending on decay rate θ_i. The active neighbors of a node affect its activation level either by excitation or inhibition, depending on the nature of the link. Each connection towards a given node i from a node j has a weight w_{ji}, which has a positive value for excitatory links and a negative value for inhibitory links (note that for each link to node i from a node j there is a reciprocal link in the network from i to j, and w_{ij} can be different from w_{ji}). At any cycle, the total input $n_i(t)$ of input from its neighbors on node i is the inner product of the vector of the neighbors' activation levels by the connection weight vector:

$$n_i(t) = \sum_j w_{ji} a_j(t)$$

This input is constrained by a "squashing function" which prevents the activation level from going beyond a defined minimum and maximum. A node becomes increasingly difficult to activate as its level approaches the maximum, and increasingly difficult to inhibit as it approaches the minimum. In mathematical terms, when the overall input is *activatory,* that is, when $n_i(t) > 0$, the actual effect on node i is determined by:

$$\varepsilon_i(t) = n_i(t)\ (\max - a_i(t)).$$

When the overall input is *inhibitory,* the actual effect is determined by:

$$\varepsilon_i(t) = n_i(t)\ (a_i(t) - \min).$$

The new activation level of a node i at time $t + \Delta t$ corresponds to the activation of this node at time t, modified by its neighbors' effect and diminished by the decay $\theta_i(a_i(t) - r_i)$:

$$a_i(t + \Delta t) = a_i(t) + \varepsilon_i(t) - \theta_i(a_i(t) - r_i).$$

Finally, each node is affected by a threshold value τ_i, below which it remains inactive.

Connection weights are fixed *a priori.* In early experiments, they were the same for all connections, but we discovered that "gang effects" appear due to extreme imbalance among words having few senses and hence few connections, as well as words containing up to 80 senses and several hundred connections, and that therefore dampening is required. In our current experiments, connection weights are determined by a simple decreasing function of the number of outgoing connections from a given node. We have applied no other criteria for assigning weights, although several possibilities, such as information about frequency, collocations, synonomy and other semantic relations, suggest themselves. However, all of these possibilities would require additional and possibly sophisticated processing to be implemented. Before we pursue these options, we would like to first determine how much can be accomplished with a network which requires only simple procedures to create, as well as the extent of the semantic information implicit in this network.

5. Testing the model

5.1 Method

We selected 23 words with which to conduct our initial experiments. Each of these words has at least two homographs (with unrelated etymologies--for example, *ash* as residue and *ash* as tree) in the *CED,* and thus at least two clearly distinguishable senses exist for each. For each homograph, one sense for which a one-to-one mapping exists between the *OALD* and the *CED* was chosen and tested. We have so far restricted our experiments to one-to-one mappings in order to test the viability of our method for the simple case, before moving on to more complex cases. Also, restricting the experiments to only the straightforward cases enables close monitoring of the network's behavior. Further, if we can handle one-to-one mappings, we can handle many-to-one mappings as well, since in these cases the network will run independently on different definitions and simply produce the same result. Means to handle other mapping situations are discussed in section 5.2.

The characteristics of the *OALD* and *CED* definitions for our corpus are described in figure 4. In computing the number of senses for a given word, we assume that semicolons within definition texts indicate a division between senses, and we disregard senses which apply to compounds and derived forms. In general, we test one sense per homograph for each word in the corpus, but note that only one sense of *sage* was tested since no staightforward mapping for an *OALD* sense of *sage 2* exists. We also tested only one sense of *sash*, because the entry for *sash 2* in the *OALD* contains only a sense for the compound *sash-window.*

To achieve sense mapping, we apply the spreading activation algorithm described in section 4 over a network created from definition texts in the *CED.* Input to the network consists of a headword together with the words in one of its sense definitions from the *OALD.* This sense is to be mapped to a sense in the *CED* for the input headword. The *OALD* definition texts are pre-processed in the same way as the definitions used to create the network, as described in section 3. After the spreading activation algorithm is run, only one sense node attached to the input headword node is active in the network. This node should identify the *CED* sense that corresponds to the *OALD* input sense.

#	word	OA		Cl	
		hom	senses	hom	senses
1	*ash*	2	4	3	8
2	*bay*	5	15	5	22
3	*cape*	2	3	2	3
4	*colon*	2	2	4	7
5	*dam*	2	6	4	7
6	*fluke*	3	5	3	8
7	*lawn*	2	3	2	3
8	*mead*	2	2	2	2
9	*mull*	3	3	5	5
10	*pike*	4	6	4	7
11	*pitcher*	2	3	2	5
12	*poker*	2	2	2	3
13	*port*	5	9	6	14
14	*punch*	3	9	3	13
15	*pupil*	2	2	2	3
16	*reel*	3	10	3	10
17	*rook*	3	6	2	4
18	*sage*	2	6	2	6
19	*sash*	2	1*	2	3
20	*sty*	2	2	2	4
21	*tend*	2	4	2	7
22	*tick*	4	8	4	12
23	*viola*	2	2	2	3
total		61	113	68	159
aver		2.7	4.9	3.0	6.9

Figure 4. Test Corpus Characteristics

* Note that while there are two entries for *sash* in the OALD, the second contains only a compound, *sash-window,* and has been excluded.

5.2. Results

The spreading activation procedure was applied to 59 input senses from the *OALD.* On average, an input definition contains 4.8 content words. The networks constructed from the *CED* for our experiments contain an average of 3347 word nodes and 2866 sense nodes, and an average of 6212 nodes altogether. The networks also contain an average of 2866 word-to-sense node transitions, 10761 sense-to-word node transitions, and 13627 total transitions.

The correct homograph in the *CED* was identified by the network in 57 of the 59 cases (97%). In 5 cases, the network identified the right homograph, but selected another sense within that homograph. Therefore, the correct homograph and the correct sense were identified in 53 of the 59 cases (90%). On the same corpus (using the same pre-processed input and definition texts), Lesk's strategy identifies the correct homograph in only 83% of the cases and the correct sense in 74% of the cases.

The two cases where the network identified the wrong homograph pertain to the same word, *port.* This example is particularly difficult, largely because *port* has several homographs with some similar vocabularly among them. The definitions that should match in the first failure are

(OALD)	**port**	**1.2**	town or city with a harbour, esp one where customs officers are stationed.
(CED)	**port**	**1.1**	a town or place alongside navigable water with facilities for the loading and unloading of ships.

These definitions are connected by one common word, *town,* but otherwise they not only do not share vocabularly but also focus on very different properties of a port-as-harbor. The network incorrectly maps *port 1.2* to

(CED)	**port**	**4.5**	*Chiefly Scot.* a gate or portal in a town or fortress.

which, like the correct sense, contains a connection through *town*, but *port 4.5* is favored because in addition, *fortress* is highly connected to *city.*

In the second failure, the definitions that should match are:

(OALD)	**port**	**2**	*Naut.* opening in the side of a ship for entrance, or for loading and unloading cargo.
(CED)	**port**	**4.1a**	*Nautical.* an opening in the side of a ship, fitted with a watertight door, for access to the holds.

Instead, the networks selects *port 1.1* (see above), because it is misled by the presence of *loading* and *unloading.*

Two observations about these failures can be made:

(1) Our strategy does not take into account indications concerning domain, register, geographic variants, etc. A refined version of our program should use these indications to select a sense or to restrict the possibilities for the correct sense. Both of the above failures would be avoided if this strategy were applied, since in the first case the indication *Chiefly Scottish* would eliminate *port 4.5* as a candidate, and in the second case the domain indication *Nautical* (which appears only on sense 4.1 of *port*) would select the correct sense immediately.

(2) It is apparent in examining our results closely that far greater accuracy can be obtained by splitting the input definitions into *genus* and *differentiae,* which can be accomplished with a simple

procedure (Chodorow, Byrd, and Heidorn, 1985). Currently, we treat all of the words in the input definition equally, and thus the words at the end of the definition have exactly the same impact as the words at the beginning. We could, instead, adopt a two-step stategy in which the genus term is mapped first and, where necessary, further discriminated in a second run of the network with the differentiae. Such a strategy would also eliminate both failures for *port,* since the genus term is identical in the correctly mapped senses in each case. This strategy should work even in cases where the genus terms in the input and target senses are not identical, since they must be very closely related semantically and should therefore be identified by the network.

The other incorrect mappings in our study involve cases where the wrong sense is selected within the right homograph. For example,

(OALD)	**ash 1a**	forest tree with silver-grey bark and hard, tough wood.

maps to

(CED)	**ash 2.2**	the close-grained durable wood of any of these trees, used for tool handles, etc.

instead of

(CED)	**ash 2.1**	any oleaceous tree of the genus *Fraxinus*, esp. *F. excelsior* of Europe and Asia, having compound leaves, clusters of small greenish flowers, and winged seeds.

Ash 2.2 is selected simply because it shares more words (*wood* and *tree*) with the *OALD ash 1a.* Again, a strategy which takes into account the *genus/differentiae* structure of the input definition would correct this failure because of the overlap in the genus term in the *OALD ash 1a* and the *CED ash 2.1.*.

Several other improvements to our model suggest themselves. As noted above, we iterate the process of building word-to-sense-to-word links two times in our current implementation. The results may be improved with additional iterations or with a single network covering the entire dictionary. Also, because the parameters used in these experiments are a first approximation, we are experimenting with network parameters such as the amount of feedback or inhibition, etc., to see what effect such changes could have on the results. Experiments with neural networks similar to ours show that such networks are extremely sensitive to tuning of this kind. Also, in our current model, all sense nodes for all homographs of a given word are attached directly to its word node, and therefore they are all mutually inhibited. Instead, senses within the same homograph (which are semantically related) should inhibit each other less than senses between different homographs (which are not semantically related). Therefore, we may see some improvement in the results if the model is modified to add intermediary nodes accounting for different homographs, as well as for subsenses, leading to a more complex, hierarchical unit for each word. Finally, we are beginning to explore means to vary weights on links within the network, on the basis of information extracted from other sources concerning frequency, part of speech, collocations, synonomy, etc., and semantic information extracted from the definition texts themselves.

As noted above, many-to-one mappings are simply one-to-one mappings where several senses map to the same sense in the target dictionary, and therefore they are handled by the strategy we have outlined here. It should be further noted that one-to-many mappings from dictionary *A* to dictionary *B* are just the reverse of many-to-one mappings from *B* to *A*. Therefore, we can handle one-to-many mappings from the *OALD* to the *CED* by constructing the network with the *OALD,* and activating it with definitions from the *CED.* However, with slight changes to our model it should be possible to handle one-to-many mappings directly. By reducing the inhibition between senses associated with the words in the network, it would be possible to activate more than one sense at the end of a network run. In this way, multiple associations could be revealed, and their relative strengths could be considered.

5. Conclusion

Although our model is only preliminary, the results are promising and improve on the results obtained with Lesk-based strategies. Like Lesk's method, the spreading activation strategy works well when common words appear in corresponding definitions. However, it is also successful in cases when no overlapping words are present and when the same number of overlaps exist with more than one sense definition, since additional, more remote relations are identified.

More generally, our results provide some positive evidence in the debate over whether the enormous body of lexical knowledge encoded in dictionaries can be exploited for natural language processing. The success of our method for both sense mapping and word sense disambiguation shows, more decisively than the work of Lesk and others working in the same paradigm, that the structure of dictionary definitions themselves provides usable semantic information about the relatedness of words. Further, our results give some evidence that the spreading activation strategy is one means to exploit the full potential of relations implicitly encoded in machine-readable dictionaries.

References

AMSLER, R. A. (1980). *The structure of the Merriam-Webster Pocket Dictionary.* Ph. D. Dissertation, University of Texas at Austin.

BYRD, R. J., CALZOLARI, N., CHODOROW, M. S., KLAVANS, J. L., NEFF, M. S., RIZK, O. (1987) Tools and methods for computational linguistics. *Computational Linguistics,* 13, 3/4, 219-240.

BYRD, R. J. (1989) Discovering Relationships among Word Senses. *Proceedings of the Fifth Annual Conference of the UW Centre for the New Oxford English Dictionary,* Oxford, 67-79.

CALZOLARI, N. (1984). Detecting patterns in a lexical data base. *Proceedings of the 10th International Conference on Computational Linguistics, COLING'84,* Stanford, 170-173.

CHODOROW, M. S., BYRD. R. J., HEIDORN, G. E. (1985). Extracting semantic hierarchies from a large on-line dictionary. *Proceedings of the 23rd Annual Conference of the Association for Computational Linguistics,* Chicago, 299-304.

CHODOROW, M. S., RAVIN, Y. , SACHAR, H. E. (1988) A tool for investigating the synonomy relation in a sense disambiguated thesaurus. *Proceedings of the 2nd ACL Conference on Applied Natural Language Processing,* 144-51.

IDE, N.M., VERONIS, J. (1990a). Refining taxonomies extracted from machine-readable dictionaries, *ALLC/ACH'90 Conference,* Siegen, Germany [to appear in selected proceedings, Oxford University Press]

IDE N.M., VERONIS J. (1990b). Very large neural networks for word-sense disambiguation. *9th European Conference on Artificial Intelligence, ECAI'90,* Stockholm.

LESK, M. (1986). Automated Sense Disambiguation Using Machine-readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. *Proceeedings of 1986 SIGDOC Conference.*

MARKOWITZ, J., AHLSWEDE, T., EVENS, M. (1986). Semantically significant patterns in dictionary definitions. *Proceedings of the 24rd Annual Conference of the Association for Computational Linguistics,* New York, 112-119.

MCCLELLAND, J. L., RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review,* 88, 375-407.

SLATOR, B.M., WILKS. Y. (1987). Towards Semantic Structures for Dictionary Entries. *Proceedings of the 1987 Rocky Mountain Conference on Artificial Intelligence,* Boulder, CO, 85-98.

VERONIS J., WURBEL, N., HARIE, S., IDE, N.M. (1990). Contruction et exploitation d'une base de données lexicale multi-dictionnaires. *10th International Workshop on Expert Systems and their Applications,* Avignon, 85-104.

VERONIS, J., IDE, N.M. (1990). Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries. *13th International Conference on Computational Linguistics, COLING'90,* Helsinki, vol. 2, 389-94.