A LEXICON OF DISTRIBUTED NOUN REPRESENTATIONS CONSTRUCTED BY TAXONOMIC TRAVERSAL

Richard F.E. Sutcliffe[1], Donie O'Sullivan, Fergus Meharg

Department of Computer Science and Information Systems
University of Limerick, Ireland

1 INTRODUCTION

In order to construct systems which can process natural language in a sophisticated fashion it is highly desirable to be able to represent linguistic meanings in a computationally tractable fashion. One approach to the problem of capturing meanings at the lexical level is to use a form of distributed representation where each word meaning is converted into a point in an n-dimensional space (Sutcliffe, 1992a). Such representations can capture a wide variety of word meanings within the same formalism. In addition they can be used within distributed representations for capturing higher level information such as that expressed by sentences (Sutcliffe, 1991a). Moreover, they can be scaled to suit a particular tradeoff of specificity and memory usage (Sutcliffe, 1991b). Finally, distributed representations can be processed conveniently by vector processing methods or connectionist algorithms and can be used either as part of a symbolic system (Sutcliffe, 1992b) or within a connectionist architecture (Sutcliffe, 1988). In previous work we have shown how such representations can be constructed automatically by the method of taxonomic traversal, using the Merriam Webster Compact Electronic dictionary (Sutcliffe, 1993) and the Irish-Irish An Foclóir Beag (Sutcliffe, McElligott and Ó Néill, 1993). However our efforts so far have been limited by our parsing technology to lexicons of a few thousand words. We describe here how we can generate a lexical entry for any of the 71,000 nouns[2] in the Princeton WordNet (Beckwith, Fellbaum, Gross and Miller, 1992), and the initial tests we have conducted on the representations.

Our method is closely related to other work which exploits the taxonomic nature of dictionary definitions (Amsler, 1980; Hiedorn, Byrd and Chodorow, 1986; Vossen, 1990; Guthrie, Slator, Wilks and Bruce, 1990; Nutter, Fox and Evens, 1990). In addition there have already been some very interesting approaches to the construction of distributed semantic representations either from dictionaries (Wilks et al., 1990) or from corpora (Schuetze, 1993).

2 EXTRACTING FEATURE REPRESENTATIONS

The object of our work is to produce for each noun-sense in a lexicon a semantic representation consisting of a set of feature-centrality pairs. The features are semantic attributes each of which says something about the concept being defined. The centrality associated with each feature is a real number which indicates how strongly the feature contributes to the meaning of the concept. The use of centralities allows us to distinguish between important and less important features in a semantic representation. By scaling the centralities in a particular noun-sense representation so that the sum of their squares is one we can use the dot product operation to compute the semantic similarity of a pair of concepts. A word compared to itself always scores one while a word compared to another word is always less than or equal to one. This is equivalent to saying that each word representation is a vector of length one in an n-dimensional space, where n is the number of features which are used in the lexicon as a whole.

Our algorithm for constructing the representations is based on two well-known observations. Firstly, a word definition in a dictionary provides attribute information about the concept ('a mastiff is a LARGE dog'). Secondly a word definition also provides taxonomic information about the concept ('a mastiff is a large DOG'). We use the former to derive attributes for our representation, and the latter to obtain other definitions higher up in the taxonomy from which further attributes can be obtained. In assigning centralities to features, we use the same value for each attribute added at a particular level in the taxonomic hierarchy, and we reduce the value used as we move up to higher levels. This corresponds to the intuition that a feature which is derived from a definition which is close to the word of interest in the taxonomy contributes more to its meaning than one which is derived from a more distant definition.

The Princeton WordNet is very suitable for use in implementing our extraction algorithm because taxonomic links are represented explicitly by pointers. In most MRDs such links have to be deduced by syntactic and semantic analysis of sense definitions. Nouns in WordNet are organised around *synsets*. Each synset may include a list of synonyms, pointers to hyponym and hypernym synsets, and a gloss corresponding to a conventional dictionary definition.

[2] This figure includes hyphenated terms, compound nouns and proper names.

Figure 1. Synset Hierarchy for the word 'terrier' derived from Princeton Wordnet.

```
1 sense of terrier

Sense 1
terrier --
   (any of several usu. small short-bodied breeds originally trained
    to hunt animals living underground)

 => hunting dog --
       (a dog used in hunting game)

  => domestic dog, pooch, Canis familiaris --
          (domesticated mammal prob. descended from the common wolf; occurs in
           many breeds)

   => dog

    => canine, canid --
           (any of various fissiped mammals with nonretractile claws and
            typically long muzzles)

     => carnivore --
            (terrestrial or aquatic flesh-eating mammal; terrestrial carnivores
             have four or five clawed digits on each limb)

      => placental mammal, eutherian, eutherian mammal

       => mammal --
              (any warm-blooded vertebrate that nourish their young with milk
               and having the skin more or less covered with hair; young are
               born alive except for the small subclass of monotremes)

        => vertebrate, craniate --
               (animals having a bony or cartilagenous skeleton with a
                segmented spinal column and a large brain enclosed in a skull
                or cranium)

         => chordate

          => animal, animate being, beast, brute, creature, fauna --
                 (a living organism characterized by voluntary movement)

           => life form, organism, being, living thing --
                  (any living entity)

            => entity --
                   (something having concrete existence; living or nonliving)
```

Table 1. Twenty words in five categories				
cars	dogs	flowers	trees	people
chariot	pug	pansy	larch	bruiser
motorbike	terrier	daffodil	pine	patriarch
jeep	lapdog	tulip	oak	siren
moped	chihuahua	rose	sycamore	rake

Table 2. Lexical Representation Summary	
No of words	20
Total number of features	249
Average number of features	39
Minimum	17
Maximum	76

The extraction algorithm starts with the synset corresponding to the word-sense for which we wish to create a lexical entry. The gloss is tokenised, function words are removed and the remaining content words are converted to their root inflection. All such words are considered to be features of the word-sense, and are given a centrality of 1.0. We then chain upwards using a hypernymic link (if any)[3]. At the next level up, features are extracted from the hypernym's gloss, using a centrality of 0.9. The process is repeated, reducing the centrality by 0.1 at each level, until either the top of the hierarchy is reached or the centrality falls to zero. Finally, the representation, consisting of a set of feature-centrality pairs, is normalised.

3 RESULTS

The algorithm described above has been implemented and can be used to construct a lexical entry for any of the nouns in the WordNet database. Figure 1 shows the synset hypernym hierarchy for the word 'terrier' in WordNet. Figure 2 shows the semantic representation derived by the algorithm for this word. We present here some preliminary experiments which attempt to measure the performance of the lexicon. Four words were chosen from each of five categories of noun which we label cars, dogs, flowers, trees and people. These are shown in Table 1. Table 2 shows a summary of the characteristics of the word representations in the set of twenty words. Pairs of categories were chosen, cars-dogs, flowers-trees and so on, each containing eight words. A series of eight-by-eight tables was then computed, showing the dot product of each word with every other word in the category pair. Table 3 shows the results for the cars-dogs matrix. There are several points to note about this table. Firstly, the match of one car word with another is high, ranging between 0.58 and 1.0 with an average of 0.8. This shows that the lexicon has captured the similarity between the car concepts. Secondly, the match of one dog word with another is also high, ranging between 0.63 and 1.0 with an average of 0.76, for the same reason. Thirdly, the match of a car word with a dog word is low, ranging between 0.05 and 0.17 with an average of 0.1. This is because cars and dogs are not closely linked semantically. Table 4 shows results for the flowers-trees matrix. Flowers and trees are much more closely related semantically than cars and dogs, and this is reflected in the results. Flower words match with tree words in a range of 0.30 to 0.67

---

[3] At present we only choose the first such link if there are several.

with an average of 0.4, much higher than for cars and dogs. The match of flowers with flowers or trees with trees continues to be high. Finally, Table 5 shows the people-dogs matrix. Note here that the match of people with themselves is lower than that of dogs with themselves (average 0.63 rather than average 0.76.) This is because the people words are in fact a rather disparate set. Note in particular that 'bruiser' against 'rake' is the best match while 'bruiser' against 'patriarch' is the worst. This matches one's intuitions about these concepts: patriarchs are "good" while 'bruisers' and 'rakes' are not.

4 CONCLUSIONS

We have presented a simple algorithm which allows a set of distributed lexical semantic representations to be constructed from nouns in the Princeton WordNet. The results show that the method works and produces good results. The main reason for this is the explicit taxonomic information in WordNet which has to be inferred in other dictionaries. Incorrect taxonomic information seriously degrades the performance of this kind of method. On the other hand errors in individual features are not so harmful as they have no knock-on effects. However, we are engaged in eliminating errors in word sense and syntactic category which are the principal sources of inaccuracy in the method. In addition we are working on objective methods for measuring the performance of the lexicon on a large scale.

5 REFERENCES

Amsler, R.A. (1984). Machine-Readable Dictionaries. *Annual Review of Information Science and Technology (ARIST)*, 19, 161-209..

Beckwith, R., Fellbaum, C., Gross, D., & Miller, G. A. (1992). WordNet: A Lexical Database Organised on Psycholinguistic Principles. In U. Zernik (Ed.) *Using On-line Resources to Build a Lexicon.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Evens, M. (1989). Computer-Readable Dictionaries. *Annual Review of Information Science and Technology (ARIST)*, 24, 85-117.

Guthrie, L, Slator, B.M, Wilks, Y. and Bruce, R. (1990). Is there Content in Empty Heads? In *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90), Helsinki, Finland,* 3, 138-143.

Hiedorn, Byrd and Chodorow (1986).

Lesk, M. (1986). Automated Word Sense Disambiguation using Machine-Readable Dictionaries:

Table 3. Cars vs. Dogs

	chariot	motorbike	jeep	moped	pug	terrier	lapdog	chihuahua
chariot	1.00	0.74	0.58	0.73	0.13	0.17	0.14	0.09
motorbike	0.74	1.00	0.69	1.00	0.11	0.11	0.11	0.06
jeep	0.58	0.69	1.00	0.68	0.08	0.09	0.09	0.05
moped	0.73	1.00	0.68	1.00	0.10	0.10	0.11	0.05
pug	0.13	0.11	0.08	0.10	1.00	0.68	0.65	0.69
terrier	0.17	0.11	0.09	0.10	0.68	1.00	0.63	0.72
lapdog	0.14	0.11	0.09	0.11	0.65	0.63	1.00	0.67
chihuahua	0.09	0.06	0.05	0.05	0.69	0.72	0.67	1.00

Table 4. Flowers vs. Trees

	pansy	daffodil	tulip	rose	larch	pine	oak	sycamore
pansy	1.00	0.32	0.36	0.49	0.37	0.32	0.37	0.28
daffodil	0.32	1.00	0.70	0.37	0.38	0.33	0.37	0.39
tulip	0.36	0.70	1.00	0.41	0.39	0.33	0.37	0.30
rose	0.49	0.37	0.41	1.00	0.56	0.58	0.67	0.44
larch	0.37	0.38	0.39	0.56	1.00	0.83	0.74	0.64
pine	0.32	0.33	0.33	0.58	0.83	1.00	0.83	0.62
oak	0.37	0.37	0.37	0.67	0.74	0.83	1.00	0.60
sycamore	0.28	0.39	0.30	0.44	0.64	0.62	0.60	1.00

Table 5: People vs. Dogs

	bruiser	patriarch	siren	rake	pug	terrier	lapdog	chihuahua
bruiser	1.00	0.40	0.52	0.63	0.12	0.15	0.13	0.08
patriarch	0.40	1.00	0.40	0.55	0.15	0.18	0.16	0.17
siren	0.52	0.40	1.00	0.50	0.14	0.17	0.14	0.09
rake	0.63	0.55	0.50	1.00	0.12	0.15	0.13	0.08
pug	0.12	0.15	0.14	0.12	1.00	0.68	0.65	0.69
terrier	0.15	0.18	0.17	0.15	0.68	1.00	0.63	0.72
lapdog	0.13	0.16	0.14	0.13	0.65	0.63	1.00	0.67
chihuahua	0.08	0.17	0.09	0.08	0.69	0.72	0.67	1.00

```
repn( terrier, '(any of several usu. small short-bodied breeds originally trained to
hunt animals living underground)', [[any, 0.19], [several, 0.19], [small, 0.19],
[breed,0.19], [originally, 0.19], [trained,0.19], [hunt, 0.19], [animal,1.9], [living,
0.19], [underground, 0.19], [a, 0.17], [dog, 0.17], [used, 0.17], [in, 0.17],
[hunting, 0.17], [game, 0.17], [domesticated, 0.15], [mammal, 0.15], [descend, 0.15],
[common, 0.15], [wolf, 0.15], [occur, 0.15], [many, 0.15], [various, 0.11], [fissiped,
0.11], [with, 0.11], [nonretractile, 0.11], [claw, 0.11], [typically, 0.11], [long,
0.11], [muzzle, 0.11], [terrestrial,0.096], [aquatic, 0.096], ['flesh-eating', 0.096],
[carnivore, 0.096], [have, 0.096], [four, 0.096], [five, 0.096], [clawed, 0.096],
[digit, 0.096], [on, 0.096], [each, 0.096], [limb, 0.096], ['warm-blooded', 0.057],
[vertebrate, 0.057], [nourish, 0.057], [young, 0.057], [milk, 0.057], [skin, 0.057],
[more, 0.057], [less, 0.057], [covered, 0.057], [hair, 0.057], [are, 0.057], [born,
0.057], [alive, 0.057], [except, 0.057], [monotreme, 0.057], [bony, 0.038], [skeleton,
0.038], [segment, 0.038], [spinal, 0.038], [column, 0.038], [large, 0.038], [brain,
0.038], [enclosed, 0.038], [skull, 0.038], [cranium, 0.038], [organism, 0.01],
[characterized, 0.01], [voluntary, 0.01], [movement, 0.01], [entity, 0.01], [concrete,
0.01], [existence, 0.01], [nonliving, 0.01]] ).
```

Figure 2. The semantic representation for 'terrier' produced by the algorithm.

How to Tell a Pine Cone from an Ice Cream Cone. *Proceedings of the 1986 SIGDOC Conference.*

Ide, N.M. and Veronis, J. (1990). Very Large Neural Networks for Word Sense Disambiguation. *Proceedings of the European Conference on Artificial Intelligence, ECAI'90, Stockholm, August 1990.*

Miller, G.A., Beckwith, R., Christiane Fellbaum, Gross, D. and Miller, K. (1992). Introduction to WordNet: An On-line Lexical Database. Manuscript.

Nutter, J.T., Fox, E.A. and Evens, M.W (1990). Building a Lexicon from Machine Readable Dictionaries for Improved Information Retrieval. *Literary and Linguistic Computing*, 5, 2, 129-138.

Schuetze, H. (1993). Translation by Confusion. In *Proceedings of the AAAI-93 Spring Symposium Series: Building Lexicons for Machine Translation, March 23-25, 1993, Stanford, University, CA.*

Sharkey, N.E., Day, P.A. and Sharkey, A.J.C. (1991). A Connectionist Machine Tractable Dictionary: The Very Idea. In C.-M. Guo (Ed.) Machine Readable Dictionaries. Norwood, NJ: Ablex.

Sutcliffe, R.F.E. (1988). A Parallel Distributed Processing Approach to the Representation of Knowledge for Natural Language Understanding. Unpublished doctoral thesis, University of Essex, UK.

Sutcliffe, R.F.E. (1991a). Distributed Representations in a Text Based Information Retrieval System: A New Way of Using the Vector Space Model. In *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, Chicago, Il., October 13-16, 1991.* New York, NY: ACM Press, pp. 123-132.

Sutcliffe, R.F.E. (1991b). Distributed Subsymbolic Representations for Natural Language: How many Features Do You Need? *Proceedings of the 3rd Irish Conference on Artificial Intelligence and Cognitive Science, 20-21 September 1990, University of Ulster at Jordanstown, Northern Ireland.* Berlin, FRG, Heidelberg, FRG, New York, NY: Springer Verlag.

Sutcliffe, R.F.E. (1992a). Representing Meaning using Microfeatures. In R. Reilly and N.E. Sharkey (eds) *Connectionist Approaches to Natural Language Processing.* Englewood Cliffs, NJ: Lawrence Erlbaum Associates.

Sutcliffe, R.F.E. (1992b). PELICAN: A Prototype Information Retrieval System using Distributed Propositional Representations. To Appear in H. Sorensen (Ed.) *Proceedings of AICS-91 - The Fourth Irish Conference on Artificial Intelligence and Cognitive Science, University College Cork, 19-20 September 1991.* London, UK, Berlin, FRG, Heidelberg, FGR, New York, NY: Springer-Verlag.

Sutcliffe, R. F. E. (1993). Constructing Distributed Semantic Lexical Representations using a Machine Readable Dictionary. In K.T. Ryan and R. F. E. Sutcliffe (Eds.) *Proceedings of AICS-91 - The*

Fifth Irish Conference on Artificial Intelligence and Cognitive Science, University of Limerick, 10-11 September 1992. London, UK, Berlin, FRG, Heidelberg, FGR, New York, NY: Springer-Verlag.

Sutcliffe, R. F. E., McElligott, A., & O'Neill, G. (1993). Using Distributed Patterns as Language Independent Lexical Representations. In *Proceedings of the AAAI-93 Spring Symposium Series: Building Lexicons for Machine Translation, March 23-25, 1993, Stanford, University, CA.*

Vossen, P. (1990). The End of the Chain: Where Does Decomposition of Lexical Knowledge Lead us Eventually? Esprit BRA-3030 ACQUILEX WP 010. English Department, University of Amsterdam, The Netherlands.

Wilks, Y., Fass, D., Guo, C.-M., Macdonald, J., Plate, T. and Slator, B. (1990). Providing Machine Tractable Dictionary Tools. *Machine Translation*, 5, 99-154.

Table 6 Word-Word Summary	
Cars-Cars average	0.80
Cars-Dogs average	0.10
Dogs-Dogs average	0.76
Flowers-Flowers average	0.58
Trees-Flowers average	0.40
Trees-Trees average	0.78
People-People average	0.63
Dogs-Dogs average	0.76
People-Dogs average	0.14