Using Distributed Patterns as Language Independent Lexical Representations

Richard F.E. Sutcliffe[1]
Annette McElligott
Gerry Ó Néill

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

1 Introduction

While it is possible to construct Machine Translation (MT) systems of surprising sophistication using the technique of transfer between augmented parse trees (Brockmann, 1991) few people would doubt that in order to perform a fully satisfactory translation it will ultimately be necessary to work with meaning representations. However, there are many problems with developing a computationally tractable representation scheme for linguistic meanings either at the sentential (propositional) or lexical levels. 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).

A further point of interest regarding distributed representations for nouns is that they can be extracted automatically from machine readable dictionaries. As is well known, dictionaries contain much taxonomic information (Amsler, 1984). This information can be extracted automatically by parsing the dictionary definitions and extracting taxonomic pointers from them in a recursive fashion. For example both Vossen (1990) and Guthrie et al. (1990) have constructed exhaustive taxonomies for the Longmans Dictionary of Contemporary English in this way. We have shown one manner in which it is possible to traverse such a taxonomy determining both the features germane to a particular concept and their appropriate centralities (strengths). In the present work we have taken this idea further by showing how lexicons in different languages can be mapped into the same feature space thus allowing lexical translation between those languages.

[1]We are most grateful to Tom Brehony, Colman Collins, Denis Hickey, Ken Litkowski, Tony Molloy, Redmond O'Brien and Piek Vossen for their help with this research. The authors' emails are sutcliffer@ul.ie, mcelligotta@ul.ie and oneillg@ul.ie

cealg (treachery)	iu'r (yew)
sleight 0.495074	arborvitae 0.421825
subterfuge 0.386588	larch 0.421825
shenanigan 0.281638	myrtle 0.421825
circlet 0.257248	box 0.36855
folder 0.257248	carob 0.36855
ring 0.257248	bock 0.344418
round 0.257248	cormorant 0.344418
bacillus 0.191211	tone 0.344418
atoll 0.181902	periwinkle 0.344418
beanie 0.181902	winkle 0.344418

Example System Outputs

2 The Extraction Algorithm

Our original extraction algorithm was very simple (Sutcliffe 1992c). Suppose the dictionary contains the definitions "An **Afghan** is a large dog", "A **dog** is a furry animal", and "An **animal** is a sentient entity". A representation for **Afghan** can be constructed by asserting that it is [**+large +furry +sentient**]. We opt to weight these attributes according to their distance from the concept being defined, giving: [**<large, 1.0> <furry, 0.9> <sentient, 0.8>**]. Such a representation can be considered as a vector in an n-dimensional space, where n is the total number of distinct features extracted from the dictionary by repeating the above procedure for all noun entries. Once such representations have been normalised (scaled to length one) we can determine how similar in meaning a pair of concepts are by computing, say, the dot product of their representations.

3 The Translation Method

In an initial experiment we have investigated whether it is possible to use the above extraction algorithm as a basis for lexical translation. The method was as follows. First we constructed a set of representations for 4263 nouns in the Merriam Webster Compact Electronic Dictionary in terms of a set of 2013 English features. Next we constructed a set of representations for 2928 Irish words in terms of 505 Irish features extracted from the Foclóir Beag dictionary. Thirdly we devised by hand a translation which maps each Irish feature onto one or more of the 2013 English features. Finally we converted each Irish word representation from Irish features to English features. The reason for allowing one source language feature to map to several target language features is that in many cases there was no exact translation between members of the two feature sets. In such cases we scaled the centralities of the selected target features so that their length taken as a whole was the same as the 'length' of the original source feature. Thus in translating [geal,0.601] using the English features 'bright', 'bright-colored', 'brilliant', 'light' and 'light-colored' we added the following set to the target representation: [[bright, 0.269], [bright-colored, 269], [brilliant,0.269], [light, 0.269], [light-colored, 0.269]]. The purpose here was to ensure that the overall contribution to meaning of the target English feature list was the same as that of the original source Irish feature.

The result of these stages is that we have a set of English words and a set of Irish words each of whose meanings is defined as some point in a common interlingual space defined by the English feature set. Now it is possible to translate from, say, Irish to English in the following manner. We select the Irish source word and retrieve its representation. This is then compared with the representation of each English word in the lexicon. We select as the English translation the word whose representation best matches that of the original Irish word.

4 Results and Discussion

The figure shows two example outputs of the system, translating from Irish into English. As can be seen, the system is selecting sensible translations for the source words and the results are in general good enough to encourage further work. However, there are still many problems with the system. There are two main reasons for this. Firstly, the parsers we are using can only recognise noun phrases and thus reject a large proportion of the noun definitions. As a result our lexicon for each language is only based on a part of the complete taxonomy defined by the parent dictionary. Secondly we have not as yet performed any post-processing on the distributed representations constructed from the taxonomies. For example many features are synonyms of each other, meaning that if one concept is, say, 'big' but not 'large' while another is 'large' but not 'big' the similarity between the two is not captured by the representations. We are working on algorithms which collapse such synonyms thus increasing the average frequency with which a feature occurs and thus improving the performance of the system.

As we have noted earlier, the translation between features in the source language and those of the target language is at present accomplished by hand. However a bilingual dictionary could also be used for this purpose. It should be noted that such a dictionary can be much more modest in scope than either of the monolingual dictionaries participating in any language mapping, and that errors in translation caused by, say, problems in resolving source or target word senses are much less injurious to the integrity of the translation than would be the case if a bilingual dictionary were being used for the translation per se.

In summary, we have presented a method for mapping between noun senses in pairs of monolingual dictionaries. The method involves constructing distributed representations for noun meanings in terms of the appropriate host language features, and then translating features in one language into the other. We have demonstrated the method using words from the CED and Foclóir Beag dictionaries.

5 References

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

Brockmann, D. (1991). Transfer Problems. Internal Memorandum ET IM 026, Computational Linguistics and Machine Translation Group, Department of Language and Linguistics, University of Essex, UK.

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.

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. Also available as Technical Report UL-CSIS-91-6, Department of Computer Science and Information Systems, University of Limerick, March 1991.

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. Also available as Technical Report UL-CSIS-90-6, Department of Computer Science and Information Systems, University of Limerick, April 1990.

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. Also available as Technical Report UL-CSIS-90-2, Department of Computer Science and Information Systems, University of Limerick, January 1990.

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. Also available as Technical Report UL-CSIS-91-9, Department of Computer Science and Information Systems, University of Limerick, April 1991.

Sutcliffe, R.F.E. (1992c). Constructing Distributed Semantic Lexical Representations using a Machine Readable Dictionary. To appear 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. Also available as Technical Report UL-CSIS-92-8, Department of Computer Science and Information Systems, University of Limerick, April 1992.

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.