

Musa Alkhalifa

Cognitive Science and Language; member of Arabic WordNet project, University of Pompeu Fabra, Barcelona, Barcelona, SPAIN.

musa.alkhalifa01@campus.upf.es

Arabic WordNet and Arabic Natural Language Processing**Abstract:**

This paper presents Arabic WordNet project (2005-2007) which aims at constructing a wordnet for Standard Arabic. This project is supported by the American government and headed by Princeton University. European universities participating in this project include Politechnical University of Catalonia, University of Barcelona, The University of Manchester. Consultants from Irion Technologies and Articulate Software also guide the construction of Arabic WordNet.

Currently, there are few publicly available lexical resources for Standard Arabic. The first objective of the project is to construct and develop a freely available lexical database for Standard Arabic. The construction of Arabic WordNet follows the design and methodology of the Princeton English WordNet (Fellbaum, 1998) and EuroWordNet (Vossen 1998). The semantic network structure of wordnet has proved to be a good model for human knowledge representation and an efficient approach for many Natural Language Processing applications including information retrieval, information extraction, document classification, text summarization, word sense disambiguation, and machine translation.

Arabic wordnet will be compatible with the Princeton English WordNet (version 2.0) and the dozens of wordnets in other languages mapped onto it, enabling accurate translation on the lexical level across many languages within the Global WordNet (<http://www.globalwordnet.org>).

The second objective of the project is the extension of a formal, language independent ontology (SUMO: the Suggested Upper Merged Ontology) that will act as an interlingua for all wordnets (Niles and Pease, 2001). This ontology can serve as a suite of automated tools for cross-linguistic NLP applications.

References

1. Beesley, K. (2001) Finite-State Morphological Analysis and Generation of Arabic at Xerox, ACL/EACL 2001, July 6th, Toulouse, France : 1-8.
2. Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., (2006). Introducing the Arabic WordNet Project, in *Proceedings of the Third International WordNet Conference*, Sojka, Choi, Fellbaum and Vossen eds.
3. Black, W. J., and Elkateb, S. (2004) A Prototype English-Arabic Dictionary Based on WordNet, *Proceedings of 2nd Global WordNet Conference, GWC2004*, Czech Republic, 67-74.
4. Buckwalter, T. (2002) Arabic Morphological Analysis, [Http://www.qamus.org/morphology.htm](http://www.qamus.org/morphology.htm)
5. De Roeck, A., and Al-Fares, W. (2000) A Morphologically Sensitive Clustering Algorithm for Identifying Arabic Roots *Proceedings of the 38th Annual Meeting of the ACL*, Hong Kong, 199-206
6. Dyvik, H. (2003) Translations as a semantic knowledge source: word alignment and wordnet, Section for Linguistic Studies scientific papers, University of Bergen, Norway.
7. Dyvik, H. (2002) Translations as Semantic Mirrors: From Parallel Corpus to Wordnet1. Section for Linguistic Studies scientific papers, University of Bergen.
8. Elkateb, S and Black, W. J. (2001) Towards the Design of English-Arabic Terminological Knowledge Base, *Proceedings of ACL 2000*, Toulouse, France:113-118.
9. Elkateb, S and Black, W. J. (2004) A Bilingual Dictionary with Enriched Lexical Information, *Proceedings of NEMLAR Cairo, Egypt 2004 Arabic Language Tools and Resources*: 79-84
10. Elkateb, S. (2005) Design and implementation of an English Arabic dictionary/editor. PhD thesis, The University of Manchester, United Kingdom.
11. Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., (2006). Building a WordNet for Arabic, in *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, Genoa, Italy.
12. Farreres, J. (2005) Creation of wide-coverage domain-independent ontologies. PhD thesis, Univeritat Politècnica de Catalunya.
13. Fellbaum, C. (ed.) (1998) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
14. Niles, I., and Pease, A. (2001) Towards a Standard Upper Ontology. In: *Proceedings of FOIS 2001*, Ogunquit, Maine, pp. 2-9.
15. Pease, A., (2000) Standard Upper Ontology Knowledge Interchange Format. Web document <http://suo.ieee.org/suo-kif.html>.

16. Pease, A., (2003) The Sigma Ontology Development Environment, in Working Notes of the IJCAI-2003 Workshop on Ontology and Distributed Systems. Volume 71 of CEUR Workshop Proceeding series
17. Pustejovsky, J. (1995) The Generative Lexicon, Massachusetts Institute of Technology.
18. Vossen, P. (ed.) (1998) EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Dordrecht: Kluwer Academic Publishers.
19. Vossen, P. Peters, W., J. Gonzalo. (1999). 'Towards a Universal Index of Meaning'. Proceedings of the ACL-99 Siglex workshop, University of Maryland, 81-90.
20. Vossen P. (2004) EuroWordNet: a multilingual database of autonomous and language-specific wordnets connected via an Inter-Lingual-Index. International Journal of Lexicography, Vol.17 No. 2, OUP, 161-173.