

Arabic WordNet: New Content and New Applications

Yasser Regragui

Mohammadia School
of Engineers, Mohammed V
University of Rabat, Morocco
yasserregragui
@research.emi.ac.ma

Lahsen Abouenour

Mohammadia School
of Engineers, Mohammed V
University of Rabat, Morocco
abouenour@yahoo.fr

Fettoum Krieche

Letter Faculty of Sais-Fez, Sidi Mo-
hammed Ben Abdellah University of
Fez, Morocco
fettoum.krieche@gmail.com

Karim Bouzoubaa

Mohammadia School
of Engineers, Mohammed V
University of Rabat, Morocco
Karim.bouzoubaa@emi.ac.ma

Paolo Rosso

NLE Lab, PRHLT Research
Center, Universitat Politècnica
de València, Spain
proso@dsic.upv.es

Abstract

The Arabic WordNet project has provided the Arabic Natural Language Processing (NLP) community with the first WordNet-compliant resource. It allowed new possibilities in terms of building sophisticated NLP applications related to this Semitic language. In this paper, we present the new content added to this resource, using semi-automatic techniques, and validated by Arabic native-speaker lexicographers. We also present how this content helps in the implementation of new Arabic NLP applications, especially for Question Answering (QA) systems. The obtained results show the contribution of the added content. The resource, fully transformed into the standard Lexical Markup Framework (LMF), is made available for the community.

1 Introduction

WordNets are important as lexical resources containing not only words of the targeted language but also synsets and semantic relations between them such as synonymy, meronymy and antonymy.

Synsets are groups of words that each can substitute others in a sentence without changing its general meaning. Therefore, in Natural Language Processing (NLP), various applications used this information, especially Query Expansion (QE), Information Retrieval (IR) and Question Answering (QA) systems.

Thus, the development of new WordNets targeting new languages and dialects and/or enriching existing ones, witnessed regular experiences and research works.

Arabic, as a Semitic language spoken by around 300 million people worldwide, is concerned by these experiences as well as the use of Arabic WordNet (AWN) (Felbaum 1998; Elkateb et al., 2006) in recent NLP applications such as information retrieval (Abbache et al., 2014) E-learning of Arabic (Karkar et al., 2015), semantic-based applications (Bouhriz et al., 2015), conceptual search (Al-Zoghby and Shaalan 2015), etc.

After the first release of AWN, There were many attempts to enrich its content (Al khalifa and Rodriguez 2009; Rodriguez et al., 2008a; 2008b). Nevertheless, the gap between AWN and the Arabic language as well as other similar WNs remained one of the limitations for its use. Also, some particularities of the Arabic language, including Broken Plurals (BP), has not been sufficiently addressed. Indeed, the morphological analysis of BP is not an easy task in NLP since they are irregular forms of plurals, and cannot be identified using patterns. Making these BP forms in a resource such as AWN is much helpful for the developers of NLP applications.

In previous research (Abouenour et al., 2013), we presented a new content that has been added to the AWN in order to cover more words and synsets and, therefore, enhance the usability of this resource for Arabic NLP applications.

This paper keeps on the track of this previous research by presenting the last experiments con-

ducted using the new content of AWN and the different refinements brought by manual validation made by lexicographers. This paper also presents the transformation of the new content of AWN into the Lexical Markup Framework (LMF) format in order to make this resource available for the community in the context of the Open Multilingual WordNet project (Bond and Kyonghee, 2012), providing free access to WordNets in several languages in a common format.

The paper is organized as follows: Section 2 recalls some existing works around the AWN enrichment and its use in NLP applications. Section 3 draws a synthesis of the main techniques that we used to enrich AWN. Section 4 presents the conducted experiments to show the usability of the new content as well as to validate this content. Section 5 addresses the transformation of the enriched AWN into the LMF format. Finally, Section 6 provides a conclusion of this research and highlights the future works.

2 Related works

The AWN project followed the development of WordNets for other languages, including Euro WordNet (Vossen 1998; 1999) by focusing, first, on the most common concepts and word-senses in PWN 2.0 (Fellbaum 1998). The first release was available on 2007¹ (Black et al., 2006; Elkateb et al., 2006).

The first AWN release contains 9,698 synsets, corresponding to 21,813 Modern Standard Arabic (MSA) words, and 6 different relation types (hyponymy, meronymy, instance, etc.). A later version of AWN, is also available and contains 11,269 synsets, corresponding to 23,841 words, including new Named Entities (Rodriguez et al., 2008).² This content is smaller than the PWN 2.0 and much smaller than what is expected by Arabic NLP applications.

Indeed, although various research experiences used AWN in many Arabic NLP applications, the common limitation reported in these experiences was the shortcomings of this resource in terms of the coverage of the Arabic language.

In a previous work (Abouenour et al., 2013), we made a comparison between the size of AWN

and a dictionary for MSA on one side, and between AWN and English and Spanish WNs on another side. This comparison allowed us to measure the gap to be filled in to improve the quality and usability of this resource.

The previously mentioned experiences in using AWN (Elghamry 2008; El Amine 2009; Baldwin et al., 2010; Sharaf 2009; Benajiba et al., 2008; 2009; Kreaa et al., 2014; Suhad et al., 2015) show the importance of the target community once this resource is enriched with a content that suits the size of MSA and the expectations by Arabic NLP applications.

In this direction, it is worth mentioning that the experiences reported in other languages (English, Spanish, etc.) show the opportunity to use WordNets in more sophisticated and complex applications, such as humour detection (Reyes et al., 2010).

Nevertheless, AWN remains one of the few and important resources that can support the development of Arabic NLP applications regarding the following findings that we reported in (Abouenour et al., 2013):

- The current AWN considers the most common concepts and word-senses in PWN 2.0 so that its use in a cross-language context is possible.
- Similarly to other wordnets, AWN is connected to SUMO (Suggested Upper Merged Ontology) (Niles and Pease, 2001; Niles and Pease, 2003; Black et al., 2006). A significant number of AWN synsets was, indeed, linked to their corresponding concepts in SUMO. Statistics show that 6,556 synsets in AWN (65.56% of the synsets) are linked to 659 concepts in SUMO (65.9% out of 1000 concepts). Definitions that are provided by SUMO and its related domain-specific ontologies can be of great interest, complementing the information contained in AWN (SUMO also covers the Arabic culture domain).

Despite the above advantages of the AWN project, there were just a few attempts to enrich its content. These attempts relied on existing tools and resources. Del Gratta and Nahli (2014)

¹ <http://www.globalwordnet.org/AWN/>

² In our work, we referred to the content of the first release.

proposed an enhancement of Arabic WordNet content using the PWN and AraMorph bilingual dictionary as bilingual resource.³ This attempt resulted in adding new words and synonyms.

There was also another attempt to build a WordNet for an Arabic dialect based on the content of AWN and a parallel dictionary (Cavalli-Sforza et al., 2013).

Boudabous et al., (2013) proposed a linguistic method based on two steps to enrich AWN. The first step defines morpho-lexical patterns and the second step enriches semantic relations using these patterns. The Wikipedia resource is also used in both steps.

In comparison with those attempts, our approach is particular in that: (i) it uses techniques and resources with higher confidence, (ii) it is followed by a significant validation by lexicographers, and (iii) the usability of the enriched AWN resource is proven through experiments in the context of real-world application, i.e., Arabic QA.

3 New Content for AWN

3.1 Techniques used for AWN enrichment

The AWN lexical resource and its semantic relations showed the ability to support QE, QA and other applications (Abouenour et al., 2014; Abouenour et al., 2010), giving rise to the improvement of performance in comparison to the baseline systems, respectively. Nevertheless, this resource has many coverage shortcomings that we emphasized through the theoretical and experience-based perspectives. These shortcomings affect the usability of this resource and have been the reasons behind its limited use in Arabic NLP applications. To tackle this problem, we proposed in a previous research (Abouenour et al., 2013) an enrichment of AWN by targeting three types of content needed by Arabic QA as observed in the experience-based analysis:

- Instances or NEs enrichment: since our aim was to answer questions from the Web, we were interested in integrating YAGO (Suchanek et al., 2007) entities and relations in AWN after their automatic translation and validation. This kind of dynamic information is widely

used in questions and other texts (Abouenour et al., 2013);

- Verbs and nouns enrichment: the coverage of these main Common Linguistic Categories is poor in AWN with respect to the Arabic lexicon and the coverage registered in experiments for TREC⁴ and CLEF⁵ nouns and verbs. The proposed enrichment consists in: (i) extending the list of verb senses in AWN using the translation of both English VerbNet (Kipper-Schuler 2006) and Unified Verb Index (UVI)⁶ by means of three heuristic rules already used in the EuroWordNet project and (ii) refining the hyponymy relation among AWN noun synsets using a technique based on pattern discovery and Maximal Frequent Sequences (MFS) (García-Hernández 2007) over Web snippets and starting from a list of AWN synsets seeds (Abouenour et al., 2013).
- Broken plurals enrichment: BP is among the forms of plural that are widely and specifically used in Arabic. The analyzed questions showed that the enrichment of AWN forms in terms of BP is important to apply the QE process for a higher number of questions in real-world applications, especially QA (Abouenour et al., 2013).

The content to be added in AWN was generated from semi-automatic techniques, i.e. automatic translation and MFS, and using external resources such as YAGO, Arabic VerbNet, UVI and Web snippets. Therefore, a manual validation by lexicographers was required to guarantee a high confidence content in the enriched AWN. More details about how these techniques and resources were used can be found in (Abouenour et al., 2013).

In the next sub section, we present this validation.

3.2 Content validation

The manual validation focused on the new entries related to nouns, verbs and BPs. This vali-

³

<http://www.nongnu.org/aramorph/english/dictionaries.html>

⁴ Text REtrieval Conference, <http://trec.nist.gov/data/qa.html>

⁵ Cross Language Evaluation Forum, <http://www.clef-campaign.org>

⁶ <http://verbs.colorado.edu/verb-index/index.php>

dition involved 3 lexicographers that are also Arabic native speakers.

Each lexicographer has to validate the BP part and a specific part of nouns and verbs. For each entry, the decision to make is three-fold: (i) a given word is correct or not, (ii) the word can be member of the given synset or not, and (iii) a synset has the given relation (synonymy or hyponymy) with the given synset or not. In the latter case, the lexicographers can propose the right relation if it exists between both synsets. If the given relation is correct but not obvious, the lexicographer can mention this by “Lenient synonymy” or “Lenient hyponymy” tag. For instance, the synset “وافق - waAfaqa_v1AR” (to agree) has been assigned a new member which is the word “أقر” (to adopt), while the lexicographers classified this as “Lenient Synonymy”. Figures 1, 2 and 3 show the results of this manual validation.

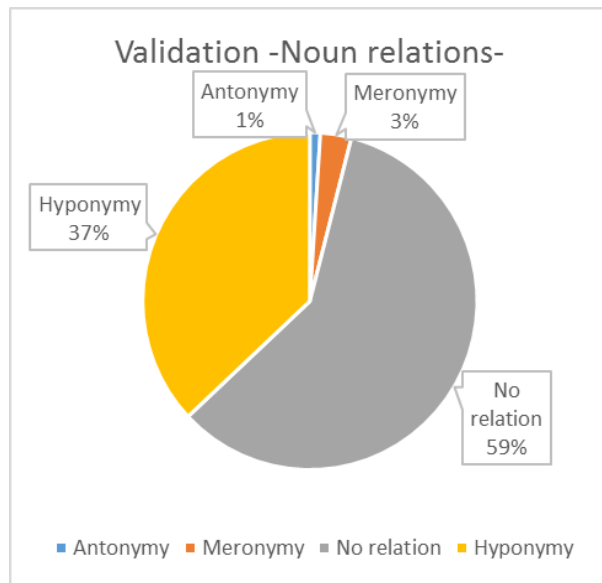


Figure 1. Manual validation of nouns

From Figure 1, 37% of the new proposed hyponymy relations between noun synsets were right. From the remaining cases, 4% were classified under the Meronymy (3%) and Antonymy (1%) relations. Hence, the overall successful relations generated by the MFS-based techniques represent 41% of the proposed ones.

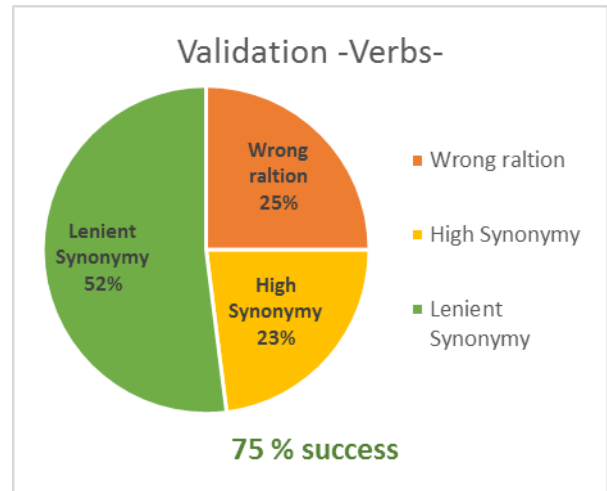


Figure 2. Manual validation of verbs

As for the verbs validation, the percentage of successful new proposed verb lemmas is roughly 75%. This percentage can be detailed as follows: (i) 23% of the new verbs can highly be re-grouped into the given synset, and (ii) 52% have lenient synonymy with the given synset members.

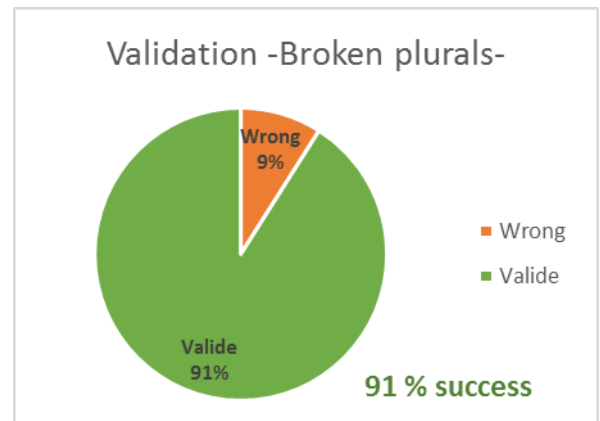


Figure 3. Manual validation of BP forms

Finally, the BP forms registered the highest percentage of success (91%), which means that the related external resource has higher precision.

4 New experiments using the enriched AWN

In (Abouenour et al., 2014; Abouenour et al., 2013; Abouenour et al., 2012), experiments were conducted in the field of Arabic QA systems using the AWN that we enriched following the techniques described above.

Let us recall that the experimental process relied on a three-level approach for Arabic QA as follows:

- Keyword-based level generating related terms from the enriched AWN;
- Structure-based level filtering passages and weighting those where question keywords and their related terms extracted from AWN appear in higher n-gram density;
- Semantic-based level comparing Conceptual Graphs (CG) of questions and candidate passages on the basis of their semantic similarity. CGs are built using AWN in addition to Arabic VerbNet.

For example, if the user question is “أين تتكون اللويحات المتسببة في مرض الزهايمر؟” (Where do plaques causing the Alzheimer's disease are made up?). The keyword-based level generates related terms of “مرض”, “المتسببة”, “اللويحات”, “الزهايمر” from AWN, the structure-based level ranks the resulting passages on the basis of the N-gram density, i.e., a passage is highly ranked when these terms and their related terms appear in it and form a high density. The semantic-based level uses the AWN and the Arabic VerbNet resources to represent the Conceptual Graph (CG) (Sowa 1983) of the resulting ranked passages as well as the question itself in order to compare the semantic similarity between both CGs (Abouenour et al., 2014). The idea of the conducted experiments is to process the three levels with the original version of AWN and thereafter with the extended version, and compare performance before and after AWN content improvement.

The effectiveness of the new content added in AWN was proven by means of different experiments previously reported in (Abouenour et al., 2013; 2014):

- **Experiment #1:** The test-set of 2,264 CLEF and TREC questions (1999-2008) shows an improvement of the keyword-based and structure-based levels using the enriched AWN according to the considered QA measures: accuracy (+53%), Mean Reciprocal Rank (+45%), answered questions (+55%) and C@1 (+50%).
- **Experiment #2:** The test-set of the 2013 QA4MRE question set (Sutcliffe et al., 2013) is composed of 284 question classified into 4 topics, namely “Aids”, “Climate change”, “Music and Society” and “Alz-

heimer”. This test-set shows the ability of the enriched AWN to support the semantic-based level at an acceptable extent (56% of the questions were represented in CG while this percentage is 61% for the candidate passages thanks to the new content of AWN).

Thus, the Experiment #1 shows that the QA performance based on the keyword-based and structure-based levels is better when using the enriched AWN. In order to show at which extent the performance of the Arabic QA process can be improved with the new AWN, we had to include the semantic-based level. The Experiment #2 shows that with the new AWN content, it is possible to semantically represent a higher percentage of questions and passages.

In this paper, we reconducted Experiment #1 using the 2,264 CLEF and TREC questions (1999-2008), considering also the semantic-based level. This is to assess the ability of the enriched AWN to improve the three levels of the Arabic QA approach regardless the considered test-set and passage collection (the QA4MRE test-set has local collection, the CLEF and TREC are assigned a Web collection).

The obtained results show a significant improvement in terms of C@1. Indeed, the three-level approach reaches a 0.51 C@1 which is higher than the 0.21 C@1 registered by the participating Arabic QA systems in QA4MRE Track 2012, including the IDRAAQ system (Abouenour 2012) that we built.

According to results of the manual validation of the new content as well as the promising results obtained with this content in a challenging task like Arabic QA, we decided to make available this content for the community of Arabic NLP.

The next section shows the steps followed to achieve this goal.,

5 The enriched AWN in LMF

5.1 Presentation of the LMF

LMF is the ISO standard for NLP lexicons and Machine Readable Dictionaries (MRD). The ISO code number for LMF is ISO-24613:2008. LMF has been developed under the aegis of TC37/SC4

by Gil Francopoulo and Monte George as editors and with Nicoletta Calzolari as convenor (Francopoulo et al., 2007).

The main goals of LMF are to use the lexical resources, manage and exchange the data among these resources, and finally, provide a common model for the creation.

Types of individual instantiations of LMF can include monolingual as in our case, bilingual or multilingual lexical resources. The same specifications are to be used for both small and large lexicons, for both simple and complex lexicons, for both written and spoken lexical representations. The descriptions range from morphology, syntax, and computational semantics to computer-assisted translation.

5.2 Transforming the raw extended AWN to LMF

As mentioned above LMF contains several packages such as syntax, morphology, semantics, MRD, and Multilingual notations. In our approach, we used the semantics and morphology packages.

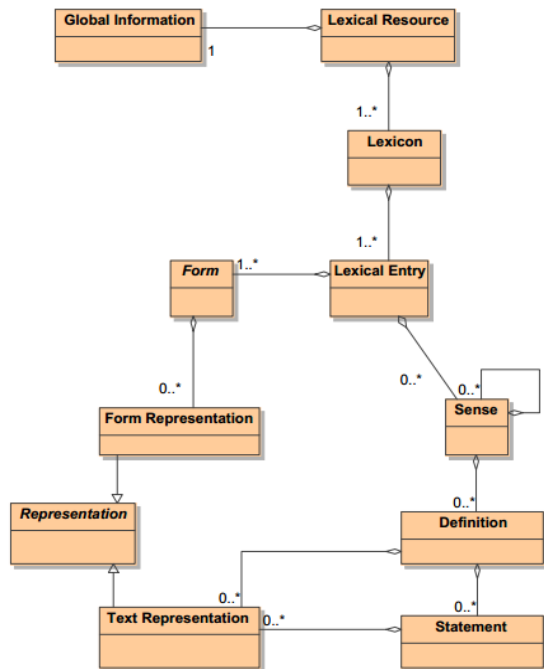


Figure 4. LMF core package.

After examining the content of the extended AWN and LMF, we made the following mapping to process the transformation:

AWN extended	LMF
Item	Sense
Word	Lexical Entry
Links	Synset relations
Forms (including BP and roots)	Word Form

Table 1. Mapping of the correspondence between LMF and the Element in the extended AWN.

Some attributes from the Element of the AWN needed to be transformed into element in the LMF to respect the DTD.

AWN extended attribute	LMF
Item id	Synset
Word value	Lemma
Form value	Written form

Table 2. List of attributes that became elements in LMF

```
<LexicalEntry id="sanap_5">
<Lemma partOfSpeech="n" writtenForm="سنة
مذنية"/>
<Sense id="sanap_5_sanap_n3AR" syn-
set="sanap_n3AR"/>
</LexicalEntry>
<LexicalEntry id="sanap_6">
<Lemma partOfSpeech="n" writtenForm="سنة
شغسية"/>
<Sense id="sanap_6_sanap_n3AR" syn-
set="sanap_n3AR"/>
</LexicalEntry>
...
<Synset baseConcept="3" id="sanap_n3AR">
<SynsetRelations>
<SynsetRelation relType="hyponym" tar-
gets="tAriyx_n1AR"/>
<SynsetRelation relType="hypernym" tar-
gets="sanap_n1AR"/>
</SynsetRelations>
</Synset>
```

Figure 5. Sample of AWN relation represented in LMF

5.3 Statistics

The process of transformation was done by a Java code using DOM API. After the transformation, we got 1036 lexical entries without part of speech (POS), this error occurred because some words did not have a corresponding item to get the POS from. To address this issue we followed a semi-automatic approach by using the SAFAR API (Jaafar and Bouzoubaa 2015; Souteh and Bouzoubaa 2011; Sidrine et al., 2010). The use of SAFAR is due to its integration of the most known Morphological Analyzers

(MA) and preprocessing tools, which simplify their use in a complementary way.

This process results in 1,012 POS from the integrated MAs, from which 24 POS were manually identified as wrong and were manually corrected.

At the end of our process, we obtained a document in LMF format totalizing 56,164 lexical entries (words grouped into synsets), 17,498 word forms and 41,136 synset relations.

6 Conclusion and Future Works

As a conclusion, the AWN project is important for Arabic NLP as witnessed by various attempts and research having used this resource. However, its content needs much extension and refinement. Our research tries to fill in a part of the gap registered between the current coverage of AWN and the expected one.

We presented in this paper the last experiments and manual validation of the AWN enrichment proposed in previous research (Abouenour et al., 2013). This enrichment was based on semi-automatic techniques and used external resources, thus the added content required refinement brought by lexicographers.

Also, the experiments conducted in the context of this paper shows a significant improvement of Arabic QA after using the enriched content of AWN. The new experiments together with the results previously presented in (Abouenour et al., 2013; 2014) show that regardless the considered test-set of questions, the new AWN content allowed an improvement of Arabic QA performance through various measures, especially the C@1 (from 0.21 to 0.51 after using the enriched AWN as a lexical resource in the keyword-based level and as a support resource for the semantic-based level).

Thus, the new release of AWN, manually validated by lexicographers and experimentally tested in the context of Arabic QA, is now available for the community in its LMF format⁷. The process of transformation into this format was described in Section 5 of this paper.

As future works, we can mention the requirement to add new relation types such as mer-

onymy and antonymy that currently are slightly present in AWN. In addition, new techniques and resources could be investigated for this enrichment.

Acknowledgement

The research of the second author was carried out in the framework of the grant provided by the Council for the Development of Social Science Research in Africa (CODESRIA) Ref. SGRT. 38/T13.

We would like to thank Dr Francis Bond, Nanyang Technological University Singapore, for his collaboration to make the new AWN v2 online for the community.

We would like to thank the lexicographers Dr Hakima Khammar, Faculty of Letters and Human Sciences, and Dr Rachida Tajmout, Mohammadia School of Engineers, for their contribution to validate the content of AWN v2.

References

- Abbache, A., Barigou, F., Belkredim, F. Z., & Belalem, G. (2014). The Use of Arabic WordNet in Arabic Information Retrieval., *International Journal of Information Retrieval Research (IJIRR)*, 4(3), 54-65.
- Abouenour, L., Bouzoubaa, K., Rosso, P. (2010). An evaluated semantic QE and structure-based approach for enhancing Arabic Q/A. In the Special Issue on Advances in Arabic Language Processing for the IEEE International Journal on Information and Communication Technologies (IJICT), ISSN: 0973-5836, Serial Publications, June 2010.
- Abouenour, L., Bouzoubaa, K., Rosso, P. (2012). ID-RAAQ: New Arabic Question Answering System Based on Query Expansion and Passage Retrieval., *CLEF'2012* (Online Working Notes/Labs/Workshop).
- Abouenour, L., Bouzoubaa, K., Rosso, P. (2013). On the Evaluation and Improvement of Arabic WordNet Coverage and Usability. In: *Languages Resources and Evaluation*, vol. 47, issue 3, pp. 891-917.
- Abouenour, L., Nasri, M, Bouzoubaa, K., Kabbaj, A., Rosso, P. (2014). Construction of an ontology for intelligent Arabic QA systems leveraging the Conceptual Graphs representation. *Journal of Intelligent and Fuzzy Systems*.
- Al Khalifa, M., & Rodríguez, H. (2009). Automatically extending NE coverage of Arabic WordNet using Wikipedia. In *Proceedings of the 3rd international conference on Arabic language processing CITALA'09*, May, Rabat, Morocco.

⁷ The AWN v2 can be downloaded from the Open Multilingual WordNet project Web site. The resource is available at: <http://compling.hss.ntu.edu.sg/omw/>

- Al-Zoghby, A. M., & Shaalan, K. (2015). Conceptual Search for Arabic Web Content. In *Computational Linguistics and Intelligent Text Processing* (pp. 405-416). Springer International Publishing.
- Baldwin, T., Pool, P., & Colowick, S. M. (2010). PanLex and LEXTRACT: Translating all words of all languages of the world. In *Proceedings of Coling 2010, Demonstration Volume*, (pp. 37-40), Beijing.
- Benajiba, Y., & Rosso, P. (2008). Arabic Named Entity Recognition using Conditional Random Fields. In: *Proc. Workshop on HLT & NLP within the Arabic world. Arabic Language and local languages processing: Status Updates and Prospects*, 6th Int. Conf. on Language Resources and Evaluation, LREC-2008, Marrakech, Morocco, May 26-31.
- Benajiba, Y., Diab M., & Rosso, P. (2009). Arabic Named Entity Recognition: A Feature-Driven Study. In: *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, num. 5. Special Issue on Processing Morphologically Rich Languages, pp. 926-934. 2009.
- Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., & Fellbaum, C. (2006). Introducing the Arabic WordNet project. In *Proceedings of the third international WordNet conference*. Sojka, Choi: Fellbaum & Vossen (eds).
- 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*, Fellbaum and Vossen (eds).
- Boudabous, M.M., Chaâben, N., Khedher, N., Hadrich Belguith, L., Sadat, F. (2013). Arabic WordNet semantic relations enrichment through morpho-lexical patterns, *The First International Conference on Communications, Signal Processing, and their Applications (ICCSPA'13)*, Sharjah, UAE, February 12-14.
- Bouhriz, N., Benabbou, F., & Benlahmer, H. (2015). Text Concepts Extraction based on Arabic WordNet and Formal Concept Analysis *International Journal of Computer Applications* (0975 – 8887) Volume 111 – No 16, February.
- Cavalli-Sforza, V., Saddiki, H., Bouzoubaa, K., Abouenour, L., Maamouri, M., & Goshey, E. (2013). Bootstrapping a wordnet for an arabic dialect from other wordnets and dictionary resources. In *Computer systems and applications (aiccsa)*, 2013 acs international conference on (pp. 1-8).
- Clark, P., & Fellbaum, C., & Hobbs, J. (2008). Using and extending WordNet to support question-answering. In: *Proceedings of the Fourth Global WordNet Conference*, University of Szeged, Hungary, pp. 111-119. COLING, pages 42.488.
- Del Gratta, R.; Nahli, O., Enhancing Arabic WordNet with the use of Princeton WordNet and a bilingual dictionary, in *Information Science and Technology (CIST)*, 2014 Third IEEE International Colloquium in , vol., no., pp.278-284, 20-22 Oct. 2014.
- El Amine, M. A. (2009). Vers une interface pour l'enrichissement des requêtes en arabe dans un système de recherche d'information. In *Proceedings of the 2nd conférence internationale sur l'informatique et ses applications (CIIA'09)*, May 3-4, Saida, Algeria.
- Elghamry, K. (2008). Using the Web in building a corpus-based hypernymy-hyponymy lexicon with hierarchical structure for Arabic. *Faculty of computers and information* (pp. 157-165).
- Elkateb, S., Black, W., Vossen, P., Farwell, D., Rodríguez, H., Pease, A., & Al khalifa, M. (2006). Arabic WordNet and the challenges of Arabic. In *Proceedings of Arabic NLP/MT conference*, London, U.K.
- Elkateb, S., Black, W., 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.
- Elkateb, S., Black, W., 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.
- Fellbaum, C. (ed.), *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press, 1998.
- Francis Bond and Kyonghee Paik (2012). A survey of wordnets and their licenses In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*. Matsue. 64-71.
- Francopoulo G., Bel N., George M., Calzolari N., Monachini M., Pet M., Soria C. 2007 *Lexical Markup Framework: ISO standard for semantic information in NLP lexicons*. GLDV (Gesellschaft für linguistische Datenverarbeitung), Tübingen
- García-Blasco, S., Danger, R., & Rosso, P. (2010). Drug-Drug interaction detection: A new approach based on maximal frequent sequences. *Sociedad Española para el Procesamiento del Lenguaje Natural, SEPLN*, 45, 263-266.
- García-Hernández, R. A. (2007). Algoritmos para el descubrimiento de patrones secuenciales maximales. Ph.D. thesis, INAOE. September, Mexico.
- García-Hernández, R. A. (2007). Algoritmos para el descubrimiento de patrones secuenciales maximales. Ph.D. thesis, INAOE. September, Mexico.
- García-Hernández, R. A., Martínez Trinidad, J. F., & Carrasco-ochoa, J. A. (2010). Finding maximal sequential patterns in text document collections and single documents. *Informatica*, 34(1), 93-101.

- Jaafar, Y., & Bouzoubaa, K. Arabic Natural Language Processing from Software Engineering to Complex Pipelines Cicing Cairo, Egypt 4/ 2015.
- Karkar, A., Alja'am, J. M., Eid, M., & Sleptchenko, A. (2015). E-LEARNING MOBILE APPLICATION FOR ARABIC LEARNERS. *Journal of Educational & Instructional Studies in the World*, 5(2).
- Kipper-Schuler, K. (2006). *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D Thesis.
- Kreaa, A., Ahmad S Ahmad and Kassem Kabalan (2014). Arabic Words Stemming approach using Arabic WordNet. *International Journal of Data Mining & Knowledge Management Process (IJDKP)* Vol.4, No.6, November 2014.
- Mousser, J. A Large Coverage Verb Lexicon For Arabic. In: *Proceedings of the 7th conference on International Language Resources and Evaluation (LREC) (2010)*, Valetta, Malta.
- Mousser, J. Classifying Arabic Verbs Using Sibling Classes. In: *Proceeding of the International Conference on Computational Semantics (IWCS) (2011)*, Oxford, UK.
- Niles, I., & Pease, A. (2001). Towards a standard upper ontology. In *Proceedings of FOIS-2* (pp. 2–9), Ogunquit, Maine.
- Niles, I., & Pease, A. (2003). Linking lexicons and ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of the 2003 international conference on information and knowledge engineering*, Las Vegas, Nevada.
- Reyes A., Rosso P., Buscaldi D. (2010). Finding Humour in the Blogosphere: The Role of WordNet Resources. In: *Proc. 5th Global WordNet Int. Conf., GWN-2010*, Bombay, India, January 31-February 4.
- Rodriguez, H., Farwell, D., Farreres, J., Bertran, M., Alkhalifa, M., & Martí, A. (2008a). Arabic WordNet: Semi-automatic extensions using Bayesian Inference. In *Proceedings of the the 6th Conference on Language Resources and Evaluation LREC2008*, May, Marrakech, Morocco.
- Rodriguez, H., Farwell, D., Farreres, J., Bertran, M., Alkhalifa, M., Martí, M., Black, W., Elkateb, S., Kirk, J., Pease, A., Vossen, P., & Fellbaum, C. (2008b). Arabic WordNet: Current state and future extensions. In *Proceedings of the fourth global WordNet conference*, January 22-25, Szeged, Hungary.
- Rodríguez, R., Farwell, D., Farreres, J., Bertran, M., Alkhalifa, M., Martí, M.A., Black, W., Elkateb, S., Kirk, J., Pease, A., Vossen, P., and Fellbaum, C., (2008). Arabic WordNet: Current State and Future Extensions. *Proceedings of The Fourth Global WordNet*
- Sharaf, A. M. (2009). The Qur'an annotation for text mining. First year transfer report. School of Computing, Leeds University. December.
- Sidrine, S., Y. Souteh, K. Bouzoubaa and T. Loukili, SAFAR: vers une Plateforme Ouverte pour le Traitement Automatique de la Langue Arabe. In: *Proceeding of the 6th Conference of Intelligent Systems: Theory and Applications SITA (2010)*, May, Rabat, Morocco.
- Souteh, Y., & Bouzoubaa, K. SAFAR platform and its morphological layer, In *Proceeding of the Eleventh Conference on Language Engineering ESOLEC'2011*, Cairo, Egypt, 14/ 12/ 2011.
- Sowa J. F. (1984). *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley Company.
- Sowa, F. *Conceptual Structures: Information Processing in Mind and Machine*, 1984, Addison-Wesley Company.
- Suchanek, F. M., Kasneci, G., Weikum, G.: YAGO: a core of semantic knowledge unifying WordNet and Wikipedia. In *Proc. of the 16th WWW*, pp. 697-706 (2007).
- Suhad A. Yousif, Venus W. Samawi, Islam Elkabani and Rached Zantout (2015). Enhancement of Arabic Text Classification Using Semantic Relations of Arabic WordNet. *Journal of Computer Science*, Volume 11, Issue 3, Pages 498-509.
- Sutcliffe, R., A. Peñas, E. Hovy, P. Forner, A. Rodrigo and C. Forascu (2013). Overview of QA4MRE Main Task, CLEF.
- Vossen P. (ed). *EuroWordNet, A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, 1999, The Netherlands.
- Vossen, P. (ed). (1998). *EuroWordNet, a multilingual database with lexical semantic networks*. Kluwer Academic Publishers, The Netherlands.