

# The Extended Arabic WordNet: a Case Study and an Evaluation Using a Word Sense Disambiguation System

**Mohamed Ali Batita**

Research Laboratory in Algebra,  
Number Theory and  
Nonlinear Analysis,  
Faculty of Science,  
Monastir, Tunisia

BatitaMohamedAli@gmail.com

**Mounir Zrigui**

Research Laboratory in Algebra,  
Number Theory and  
Nonlinear Analysis,  
Faculty of Science,  
Monastir, Tunisia

Mounir.Zrigui@fsm.rnu.tn

## Abstract

Arabic WordNet (AWN) represents one of the best-known lexical resources for the Arabic language. However, it contains various issues that affect its use in different Natural Language Processing (NLP) applications. Due to resources deficiency, the update of Arabic WordNet requires much effort. There have only been only two updates it was first published in 2006. The most significant of those being in 2013, which represented a significant development in the usability and coverage of Arabic WordNet. This paper provides a study case on the updates of the Arabic WordNet and the development of its contents. More precisely, we present the new content in terms of relations that have been added to the extended version of Arabic WordNet. We also validate and evaluate its contents at different levels. We use its different versions in a Word Sense Disambiguation system. Finally, we compare the results and evaluate them. Results show that newly added semantic relations can improve the performance of a Word Sense Disambiguation system.

## 1 Introduction

Natural language processing (NLP) is part of computer linguistics, which is also part of artificial intelligence. There are many disciplines in NLP. Information extraction is one of them. It can be text mining, information retrieval, named entity recognition. . . All these disciplines require lexical and semantic resources to proceed and generate satisfactory results. The more inclusive the resource, the more accurate the results will be. Lack of resources, especially for less-resourced language such as Arabic, has always been a persistent problem. One of

the reliable resources for the Arabic language is Arabic WordNet (AWN) (Black et al., 2006).

Princeton WordNet (PWN) (Miller, 1995; Miller, 1998), English WordNet or simply WordNet is the original and most developed of all wordnets. From its first publication, it proved its reliability with various NLP tasks. Many researchers were inspired by its usability and made a wordnet for their own languages. Now we have more than 77 wordnet<sup>1</sup>, which AWN is one. Researches now are aiming either to create new wordnets for other languages (or dialects) or improve existing ones. Creating new wordnets can be done by gathering an exhaustive repository of meanings and senses, e.g. dictionary or corpora, and assigning all words for each sense. This approach is called the merge approach (Vossen, 1998). More common is the ‘expansion’ approach. It consists of translating the core of PWN<sup>2</sup> and extending it through more concepts related to the language. This is called the top-down approach. AWN has followed this approach.

Generally speaking, a wordnet is a group of *synsets* interconnected with different relations. A synset is a set of synonyms. In other words, it is a group of words that share the same meaning. Relations can be synonymy, antonymy, hyponymy, meronymy. . . The enrichment of a wordnet can follow the axe of synsets or relations. Besides, the coverage in terms of synsets with diverse relations can be very useful in many NLP applications, especially Question Answering (QA) and Word Sense Disambiguation (WSD). Numerous approaches present themselves to construct and extend wordnets, from statistics to word embedding-based approaches (Neale, 2018).

Even without enrichment, AWN showed great results with several NLP applications like infor-

<sup>1</sup><http://globalwordnet.org/resources/wordnets-in-the-world>

<sup>2</sup>It contains the most frequently used words in any language and it has about 5,000 words.

mation retrieval (Abbache et al., 2016; Bouhriz et al., 2015) and query expansion (Abbache et al., 2018) even for e-learning applications (Karkar et al., 2015). But, AWN has seen many attempts to enrich its content with different approaches, either by adding new synsets or new entities or even new specificity of the Arabic language like broken plurals<sup>3</sup> (Abouenour et al., 2013; Saif et al., 2017; Aneur et al., 2017; Batita and Zrigui, 2017; Batita and Zrigui, 2018). Despite these efforts, AWN remains inadequate to the needs of complex modern systems. There remains a huge gap between the contents of AWN and the Arabic language itself, and also between AWN and other wordnets like PWN. This paper cites several significant programmes that have been undertaken to improve the contents of AWN. This paper also seeks to shine a light on the semantic relations of AWN and their importance for improving the performance of NLP applications. Finally, the paper provides an overview of tests we have undertaken with three versions of AWN in a concurrent NLP application.

The paper is structured as follows. The next section is an overview of the various updates and extensions of the AWN along a detailed discussion about its content. Section 3 summarises most of the significant research undertaken to enrich the semantic relations in AWN. Section 4 discusses the procedures that we follow to validate the newly added relations. Section 5 presents the conducted tests to show much the enriched AWN can affect a WSD system. Finally, section 6 will be our conclusion with some future works.

## 2 Versions of Arabic WordNet

The AWN project started in 2006. The goal was to build a freely open source lexical database for the Modern Standard Arabic available for the NLP community (Abbache et al., 2018). By that time, it has 9,698 synsets, corresponding to 21,813 words. Synsets were linked by 6 different types of semantic relations (hyponymy, meronymy, etc.), in a total of 143,715 relations (Cavalli-Sforza et al., 2013). Entities are distinguished by their part of speech POS: noun, verb, adverb, or adjective. Synsets are linked to their counterpart in PWN and the Suggested Upper Merged Ontology (SUMO) via the so-called Interlingual Index (ILI) (Black et al., 2006).

<sup>3</sup>It is non-regular plural that involves internal changing in the structure of an Arabic word.

In 2010, a second version has been published by Rodriguez et al. (Rodríguez et al., 2008). It has 11,269 synsets corresponding to 23,481 words with 22 types of semantic relationships in a total of 161,705 relations. This version has a browser written with JAVA that has an update and search functions (Rodríguez et al., 2008). This version is rich with more specific concepts related to the Arabic cultures like named entities and the Arabic language like broken plurals (Batita and Zrigui, 2018). Several researchers have taken advantage of this version in most of their work in different areas of NLP to improve the performance of their systems.

Recently, an extended version has been published in 2015 by Rezagui et al. Rezagui et al. (Rezagui et al., 2016). This version is seen as an improvement of the coverage and usability of the previous version of AWN (Abouenour et al., 2013). It includes 8,550 synsets which correspond with 60,157 words, among which we find 37,342 lemmas, 2,650 broken plurals, and 14,683 verbal roots. Rezagui et al. (Rezagui et al., 2016) changed the structure of the database to the Lexical markup framework (LMF) (Francopoulo et al., 2006), the ISO standard for NLP and machine-readable dictionary (MRD) lexicons. They made it publicly available and ready to use from the Open Multilingual Wordnet<sup>4</sup>.

Table 1 below summarizes the statistics of entities, synsets, and relations of PWN and the three previous versions of AWN.

|                  | PWN                   | V1                   | V2                    | Ex.V                |
|------------------|-----------------------|----------------------|-----------------------|---------------------|
| <b>Entities</b>  | 206,978               | 21,813               | 23,481                | 60,157              |
| <b>Synsets</b>   | 117,659               | 9,698                | 11,269                | 8,550               |
| <b>Relations</b> | 283,600<br>(22 types) | 143,715<br>(6 types) | 161,705<br>(22 types) | 41,136<br>(5 types) |

Table 1: Statistics of PWN with 3 versions of AWN.

First of all, we notice that the number of entities and synsets in PWN is very high compared to all the versions of AWN. In versions 1 and 2 (V1 and V2), we find that the number of entities is proportional to the number of synsets which is approximately two to three times the number of entities, which is not the case in the extended version (Ex.V). On the one hand, V2 contains more

<sup>4</sup><http://compling.hss.ntu.edu.sg/omw/>

synsets and fewer entities than the Ex.V. On the other hand, V2 has 11,269 synsets connected with 161,705 relations and Ex.V has only 8,550 synsets connected with only 41,136 relations. By comparing the number of relations in PWN with V2, we note that V2 is nearly rich in terms of connections between synsets. As a result, we can say that Ex.V is more affluent than the other versions of AWN in terms of synsets but impoverished in terms of relations. Abouenour et Al. (Abouenour et al., 2013) put a focus on the entities, in this paper, we focus on the relations between them.

### 3 Related Works

Until now, there are several attempts to enrich the AWN using different methods and approaches. Most of the works focused on the improvement of the number of entities and synsets (Rodríguez et al., 2008; Alkhalifa and Rodríguez, 2009; Abouenour et al., 2010; Abouenour et al., 2013; Regragui et al., 2016; Ameer et al., 2017; Saif et al., 2017; Lachichi et al., 2018). The main reason behind those works is the richness of the Arabic Language. One study on both Arabic and English Gigaword corpus has shown that to deal with the same linguistic content of 100,000 words in English, it takes approximately 175,600 words in Arabic (Alotaiby et al., 2014). In other words, one English word can be processed with approximately two Arabic words. Thus, resource-based applications expect more coverage of the Arabic language.

In contrast, the work on the relations of AWN is much less. Boudabous et Al. (Boudabous et al., 2013) proposed a linguistic method based on two phases. The first one defines morpho-lexical patterns using a corpus developed from Arabic Wikipedia. The second one uses the patterns to extract new semantic relations from the entities in AWN. A linguistic expert has validated the obtained relations. While some of the new relations were good others were not - for various reasons, including the size of the corpus and the patterns applications.

In our first work on the AWN (Batita and Zrigui, 2017) we focused on the enrichment of antonym relations. As many studies have shown that the antonym relation is universal, but, it has been noted that there are different perspectives towards this lexical relation in different cultures (Hsu, 2015). Antonyms detection, in general, is a tough task for the NLP community. After a deep study, we

have found that the extended version of AWN has only four types of relations. One of them is the antonym relations with only 14 pairs. This work has been concentrated on the extended version of AWN because it has been proved by Abouenour et al. (Abouenour et al., 2013) that it has given excellent results when testing in a Q/A system. We proposed a pattern-based approach to extract new antonym relations from the entities of AWN. For that, they extract patterns from an Arabic corpus and used a corpus analysis tool to recognize automatically the antonym pairs from other pairs. The analysis tool is the Sketch Engine (Kilgarriff et al., 2004). It has many useful metrics like the LogDice which gives a higher score to most likely related pairs. The results were filtered using the LogDice and the validation was manual.

After that our next step was the derivational relations in AWN (Batita and Zrigui, 2018). By that, we tackled another matter of the Arabic language which is the morphological aspect. The derivational and morphological problem has been a subject in different wordnet from other languages (Koeva et al., 2008; Mititelu, 2012; Šojat et al., 2012). Generally speaking, and when it comes to studying a language aspect, rule-based approaches seem the more promoting one because they rely on linguistic rules verified by an expert or by a native speaker. Based on that, we relied on that kind of approach to add new derivational relations between entities in AWN. We studied the derivational aspect of the Arabic language to make a set of transformation rules. Those rules are based on the POS switch, for example between the verb *كَتَبَ* *kataba*<sup>5</sup> (write)

and the noun *كَاتِبٌ* *kaatibun* (writer) there is a *Has-DerivedVerb* relation. Rules are made by an expert and validated carefully to guaranty the precision of the results. For more information on the transformation rules see (Batita and Zrigui, 2018). In the end, we got 8 different relations with different frequencies. The validation of the rules and the final results has been made by a lexicographer.

The knowledge-based systems in general and wordnet-based systems specifically shown good results when they used a rich wordnet with as many relations as possible (Fragos et al., 2003; Seo et al., 2004; Alkhatlan et al., 2018). Yet, the use of a wordnet, in general, has shown a great result in different areas of NLP such as humor detection

<sup>5</sup>We used the transliteration system of L<sup>A</sup>T<sub>E</sub>X.

(Barbieri and Saggion, 2014) and human feelings (Siddharthan et al., 2018) even in the cybercrime investigation (Iqbal et al., 2019). Given a sufficiently large database with many words and connections between them, many applications are quite capable of performing sophisticated semantic tasks. That is why work on the relations in AWN has to increase because richer resource can achieve significant results in a real-world NLP application. Evaluation and validation of the relations need to be considered as essential and continuous steps to guaranty the credibility of a resource. Basically, validation can be done either manually by verifying each relations individually or automatically using different approaches. In the next, we will describe how we validated the newly added relations in the previous updates.

#### 4 Validation of the New Relations in Arabic WordNet

The previously cited works on the enrichment of the relations in AWN confronted different parts of the Arabic language, in general, using different methods and approaches. Table 2 summaries all the relations (new and pre-existing) of the extended version of AWN along with their frequency.

| Relation          | Frequency     |
|-------------------|---------------|
| Hyponym           | 21,851        |
| Hypernym          | 21,851        |
| NearSynonym       | 673           |
| HasInstance       | 1,295         |
| IsInstance        | 1,295         |
| Antonym           | 800           |
| HasDerivedVerb    | 2,005         |
| ActiveParticiple  | 1,347         |
| PassiveParticiple | 1,004         |
| Location          | 985           |
| Time              | 752           |
| Instrument        | 184           |
| HasDerivedNoun    | 1,784         |
| Relatedness       | 804           |
| <b>Total</b>      | <b>56,630</b> |

Table 2: Relations of the extended version of Arabic WordNet with their frequencies.

We will focus on the extended version published by Rezagui et al. (Rezagui et al., 2016) and the new relations that we already added (Batita and Zrigui, 2017; Batita and Zrigui, 2018). Since many relations need to be validated (12), we initially

used an automatic approach, which we developed. While the majority of the new relations are specific to the Arabic language (8 derivational relations), with the developed approach we will be working only on the three general relations: hyponyms, hypernyms, hasInstance, isInstance, and synonyms. We were inspired by the aspect of the dictionary and the construction of wordnets since they are based on the synonyms and the *is-a* relations (hyponym/hypernym).

Our automatic approach says that ‘if a word  $w$  has a dictionary definition and belongs to a synset  $s$  with other words  $w_1, \dots, w_k$  then there is a strong probability that  $w$  mentions one or more of  $w_k$  in her definition and/or other words ( $w_k$ ) from the synonym/hypernymy/instance of  $s$ ’. An example will simplify the point of the view:

- $W$ : تلف *tlf* (dammage)
- $S = ta|kala\_v1AR$ : صدأ ، تلف ، تأكل، *tākl, tlf, ṣḍa* (corrosion, damage, rust)
- *Hyponym* = *AinohaAra\_v1AR*: أنهار، تدهور، فسد *āinhār, tdhwr, fsad* (collapsed, deteriorated, ruined)
- *Definition of w*: تلف الزرع، فسد، عطب *tlf ālẓrʿ, fsad, ʿṭb* (The implant is damaged, corrupted, damaged)

As we can see,  $W \in S$  and its definition have a word (فسد *fsd*) that refers to the *hyponym* of  $s$ . If so, then the relation is validated, otherwise it should be reviewed. We collect all the definition of the words that have one of the three relations from different dictionary<sup>6</sup>. All definitions are stored in one file. The file is structured as a table and each line contains one definition per word. Stop words are eliminated and remaining words have been lemmatized<sup>7</sup>. Finally, we applied our idea and we got the results of each relation as described in table 3. The high accuracy of the synonyms due to their limited number (we have only 412 relations). False relations are due to one of the following reasons (i) either a problem with the lemmatization or (ii) the granulate of the definition or (iii) the diacritization and/or correct written form of the word.

As a start-up, the first approach yields to promoting accuracy. To guaranty efficiency and high confidentiality, a second validation is done manually by native speakers and a linguistic expert. The

<sup>6</sup>For that we used the website of AlMaany <https://www.almaany.com/>.

<sup>7</sup>We used the Farasa toolkit (Abdelali et al., 2016).

| Relations              | Accuracy (%) |
|------------------------|--------------|
| HasInstance/IsInstance | 89,1         |
| Hypernym/Hyponym       | 86,2         |
| Synonym                | 96,7         |

Table 3: Accuracy of the automatic validation according to each relation.

remaining relations (derivational and wrongly validated by the first method) have been reviewed one by one. Native speakers made suggestions for some relations that may or may not hold between words. As an example, the two words *وطن* *wṭn* (prepare to do) and *نظم* *nzm* (organize) are connected by the *hyponym* relation. Native speakers suggested that it should be eliminated but the expert said otherwise. So, the expert takes the final decisions. If a relation is obvious and does not exist, the expert can add it, as well as he can eliminate it otherwise. Besides his knowledge, the decisions of the expert are based on the following conditions:

- The suggestions of the native speakers.
- A clear definition of the words in the Arabic dictionary *لسان العرب* *lsān ālʿrb* (Lisan al-Arab).
- The existence of the relation between the words in question in AWN (some words do not have any relation at all).
- The correctness of two words that hold the relation.
- The existence of a relatedness between the words in the Arabic dictionary.

In the end, we got 81% correct relations, 5% wrong relations, 12% partially wrong relations (one of the pair of the words is wrong), and 2% of the words with no relations at all. Most of the wrong relations were found in the relations that are specific to the Arabic language, like *Instrument* and *Relatedness* because they are based on transformation rules. Sometimes, words (irregular ones) that share this kind of relations do not follow any transformation rules. Some changes have been made by the linguistic expert regarding the 12% of the relations that are partially wrong by either changing one of the two words or replacing if the word does not exist in AWN. Finally, we could not do anything for the 2% of the words that have no relations at all.

## 5 Evaluation with a Word Sense Disambiguation System

In literature, we find different approaches to evaluate any lexical resources and the choice between them depending mainly on the kind of the resource itself and for what purpose (Brank et al., 2005). Since AWN is a lexical database in the first place, then its evaluation should follow one of the following strategy:

- Comparing it to a golden standard wordnet (in most cases, PWN).
- Using it in real-life NLP application and evaluating the obtained results.

As for the first approach of evaluation, many researchers have faced difficulties with it. Abouenour et al. (Abouenour et al., 2013) compared the content of AWN with the content of PWN and the Spanish WordNet. They found that the number of synsets in AWN is around 8% (too low) of those of PWN, while the Spanish wordnet represents 49%. Taghizadeh et al. (Taghizadeh and Faili, 2016), also, compared their newly constructed Persian WordNet with FarsNet and they found a precision of 19%, which is too low to consider their resource as a reliable one.

Basically, one can tell if a wordnet is a reliable resource or not by how far it can help a system to achieve better results. This kind of evaluation seems to be a better way to test the extended AWN. As mentioned above (section 2), many researchers used the AWN in their applications and it helped achieve great results. As we are concentrated on the relations of AWN, we looked into some NLP applications to see how the relations between the entities in AWN can affect the precision of an NLP application.

Word Sense Disambiguation WSD seemed the most successful system to show the effectiveness of the relations between the words. The choice of the WSD system was made following a study of different systems that profit from the relations in AWN. The aspect of the disambiguation is based on the similarity between words, which is exactly what the relations in AWN are made for in the first place. Besides, many WSD systems have been based on the relationship between words (Fragos et al., 2003; McCarthy, 2006; Kolte and Bhirud, 2009; Zouaghi et al., 2011; Zouaghi et al., 2012; Dhungana et al., 2015) and other applications, like information

retrieval and Q/A system, rely more on the words themselves rather than the relations between them. All of this gives the WSD the advantage to be our best candidate.

Since our aim is to evaluate the impact of the relations in AWN on a WSD system, the choice of the WSD algorithm is not the main task. We implement the very simple algorithm of Galley et al. (Michel and Kathleen R., 2003) with a slight difference. The algorithm proceeds as follows:

1. Build a representation of all possible combination of the text.
2. Disambiguate all words in the text.
3. Build a lexical chains.

The algorithm takes a text as an input and proceeds all of the possible combinations between the current word and all the previous words. After that, a weighted edge takes the place if one of the senses of the current word has a semantic relation with any senses of the previous words. At the end of the text, a *disambiguation graph* is built with the nodes represent the senses of each word of the text and the edges representing the semantic relations between the senses of the words since AWN links the senses and not the words. Finally, the weights of each edge are summed up to represent a final score to each sense for each word in the text. The correct sense of the target word have the highest score. One thing to mention here is that this algorithm works with only 4 semantic relations (synonym, hypernym/hyponym, and sibling) and the weight of each edge is assigned according to the type of relations and the distance between the two words.

We use the Khaleej-2004 corpus (Abbas et al., 2011). It contains 5690 documents divided to 4 categories; international and local news, economy, and sports. It has a total of nearly 3 millions words. We did not work on optimizing the weight nor the distance between the words. The only difference that we made is the number of relations. We implemented this algorithm to work with more relations. All relations in the extended version of AWN are taken into consideration. We tested the algorithm with three versions of AWN; the version 2, the extended version with and without the new relations. Table 4 shows the obtained results.

As we can see from table 4, the enriched AWN with the semantic relations yields a significant improvement with a 78,6% of precision. We remark

| Tested versions of AWN     | Precision (%) | Recall (%)  | F1 score    |
|----------------------------|---------------|-------------|-------------|
| V2                         | 69,2          | 57,6        | 72          |
| Ex.V without new relations | 72,7          | 66,9        | 69,6        |
| Ex.V with new relations    | <b>78,6</b>   | <b>71,1</b> | <b>74,6</b> |

Table 4: Precision, recall, and f1 score with different versions of AWN.

that the precision of V2 and the Ex.V without the new relations are very close. That is due to the diversity of the first one in terms of relations (22 types) and the richness of the second one in terms of hyponym/hypernym relations (19,806 relations). Despite the fact that V2 has more relations than Ex.V (161,705 and 50,787), the difference between their precisions is that V2 does not have much of specific relations related to the Arabic language. As an example, *عزف* *ʿzf* is a polysemous verb. Two of his senses are completely different. One could be ‘playing music’ and the other ‘strike.’ In the extended AWN and without the enrichment of the relations, it has only two relations, *hyponym* with the verb *شغّل* *šġl* (fill) and *hypernym* with the verb *أخرج* *ahrġ* (get it out). When we run the test in the WSD system, we could get the appropriate sense. After the test with the new relations, we got the *Instrument* relation with a higher score.

The obtained results with the enriched AWN showed the importance of the resource and the relations between its words, even in a simple knowledge-based WSD algorithm like the one we used.

## 6 Conclusion

In this paper, we presented the different versions of the AWN along with a study case on the newly added relations to its extended version. Next, we described the content of different versions of AWN with some remarkable works done to enrich its relations. Then, we cited many evaluation approaches in general and how we evaluated AWN specifically. We provided an automatic method to validate some of the relations in AWN. In the end, we found the most reliable approach is the human evalua-

tion, despite the fact that it does not take advantage of computer programs and relies heavily on time-consuming work. To make the new content more accurate, we tested different versions of AWN with a real-life NLP application (WSD system). We attended interesting and promising results with the extended version of AWN. Before making it online and ready for the NLP community, we are still working on improving and refining the semantic relations in AWN to get more accuracy and we are running some test in different NLP applications.

## References

- Ahmed Abbache, Fatima Barigou, Fatma Zohra Belkredim, and Ghalem Belalem. 2016. The use of arabic wordnet in arabic information retrieval. In *Business Intelligence: Concepts, Methodologies, Tools, and Applications*, pages 773–783. IGI Global.
- Ahmed Abbache, Farid Meziane, Ghalem Belalem, and Fatma Zohra Belkredim. 2018. Arabic query expansion using wordnet and association rules. In *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications*, pages 1239–1254. IGI Global.
- Mourad Abbas, Kamel Smaïli, and Daoud Berkani. 2011. Evaluation of topic identification methods on arabic corpora. *JDIM*, 9(5):185–192.
- Ahmed Abdelali, Kareem Darwish, Nadir Durrani, and Hamdy Mubarak. 2016. Farasa: A fast and furious segmenter for arabic. In *The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 11–16.
- Lahsen Abouenour, Karim Bouzoubaa, and Paolo Rosso. 2010. Using the yago ontology as a resource for the enrichment of named entities in arabic wordnet. In *Proceedings of The Seventh International Conference on Language Resources and Evaluation (LREC 2010) Workshop on Language Resources and Human Language Technology for Semitic Languages*, pages 27–31.
- Lahsen Abouenour, Karim Bouzoubaa, and Paolo Rosso. 2013. On the evaluation and improvement of arabic wordnet coverage and usability. *Language resources and evaluation*, 47(3):891–917.
- Musa Alkhalifa and Horacio Rodríguez. 2009. Automatically extending the coverage of arabic wordnet using wikipedia. In *Proc. Of the 3rd International Conference on Arabic Language Processing CITALA2009, Rabat, Morocco*.
- Ali Alkhatlan, Jugal Kalita, and Ahmed Alhaddad. 2018. Word sense disambiguation for arabic exploiting arabic wordnet and word embedding. *Procedia computer science*, 142:50–60.
- Fahad Alotaiby, Salah Foda, and Ibrahim Alkharashi. 2014. Arabic vs. english: Comparative statistical study. *Arabian Journal for Science and Engineering*, 39(2):809–820.
- Mohamed Seghir Hadj Ameer, Ahlem Chérifa Khadir, and Ahmed Guessoum. 2017. An automatic approach for wordnet enrichment applied to arabic wordnet. In *International Conference on Arabic Language Processing*, pages 3–18. Springer.
- Francesco Barbieri and Horacio Saggion. 2014. Automatic detection of irony and humour in twitter. In *ICCC*, pages 155–162.
- Mohamed Ali Batita and Mounir Zrigui. 2017. The enrichment of arabic wordnet antonym relations. In *International Conference on Computational Linguistics and Intelligent Text Processing*, pages 342–353. Springer.
- Mohamed Ali Batita and Mounir Zrigui. 2018. Derivational relations in arabic wordnet. In *The 9th Global WordNet Conference GWC*, pages 137–144.
- William Black, Sabri Elkateb, Horacio Rodriguez, Musa Alkhalifa, Piek Vossen, Adam Pease, and Christiane Fellbaum. 2006. Introducing the arabic wordnet project. In *Proceedings of the third international WordNet conference*, pages 295–300. Citeseer.
- Mohamed Mahdi Boudabous, Nouha Chaâben Kamoun, Nacef Khedher, Lamia Hadrich Belguith, and Fatiha Sadat. 2013. Arabic wordnet semantic relations enrichment through morpho-lexical patterns. In *2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, pages 1–6. IEEE.
- Nadia Bouhriz, Faouzia Benabbou, and Habib Benlahmer. 2015. Text concept extraction based on arabic wordnet and formal concept analysis. *International Journal of Computer Applications*, 111(16):30–34.
- Janez Brank, Marko Grobelnik, and Dunja Mladenic. 2005. A survey of ontology evaluation techniques. In *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)*, pages 166–170. Citeseer Ljubljana, Slovenia.
- Violetta Cavalli-Sforza, Hind Saddiki, Karim Bouzoubaa, Lahsen Abouenour, Mohamed Maamouri, and Emily Goshey. 2013. Bootstrapping a wordnet for an arabic dialect from other wordnets and dictionary resources. In *2013 ACS International Conference on Computer Systems and Applications (AICCSA)*, pages 1–8. IEEE.
- Udaya Raj Dhungana, Subarna Shakya, Kabita Baral, and Bharat Sharma. 2015. Word sense disambiguation using wsd specific wordnet of polysemy words. In *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*, pages 148–152. IEEE.

- Kostas Fragos, Yannis Maistros, and Christos Skourlas. 2003. Word sense disambiguation using wordnet relations. In *First Balkan Conference in Informatics, Thessaloniki*.
- Gil Francopoulo, Monte George, Nicoletta Calzolari, Monica Monachini, Nuria Bel, Mandy Pet, and Claudia Soria. 2006. Lexical markup framework (lmf). In *International Conference on Language Resources and Evaluation-LREC 2006*.
- Chan-Chia Hsu. 2015. A syntagmatic analysis of antonym co-occurrences in chinese: contrastive constructions and co-occurrence sequences. *Corpora*, 10(1):47–82.
- Farkhund Iqbal, Benjamin CM Fung, Mourad Deb-babi, Rabia Batool, and Andrew Marrington. 2019. Wordnet-based criminal networks mining for cyber-crime investigation. *IEEE Access*.
- Abdelghani Karkar, Jihad Mohamad Alja'am, Mohamad Eid, and Andrei Sleptchenko. 2015. E-learning mobile application for arabic learners. *Journal of Educational & Instructional Studies in the World*, 5(2).
- Adam Kilgarriff, Pavel Rychly, Pavel Smrz, and David Tugwell. 2004. Itri-04-08 the sketch engine. *Information Technology*, 105:116.
- Svetla Koeva, Cvetana Krstev, and Duško Vitas. 2008. Morpho-semantic relations in wordnet—a case study for two slavic languages. In *Global wordnet conference*, pages 239–253. University of Szeged, Department of Informatics.
- SG Kolte and SG Bhirud. 2009. Wordnet: a knowledge source for word sense disambiguation. *International Journal of Recent Trends in Engineering*, 2(4).
- Cilia Lachichi, Chahrazad Bendiaf, Lamia Berkani, and Ahmed Guessoum. 2018. An arabic wordnet enrichment approach using machine translation and external linguistic resources. In *2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*, pages 1–6. IEEE.
- Diana McCarthy. 2006. Relating wordnet senses for word sense disambiguation. In *Proceedings of the Workshop on Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together*.
- Galley Michel and McKeown Kathleen R. 2003. Improving word sense disambiguation in lexical chaining. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, pages 1486–1488.
- George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- George Miller. 1998. *WordNet: An electronic lexical database*. MIT press.
- Verginica Barbu Mititelu. 2012. Adding morpho-semantic relations to the romanian wordnet. In *LREC*, pages 2596–2601.
- Steven Neale. 2018. A survey on automatically-constructed wordnets and their evaluation: Lexical and word embedding-based approaches. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.
- Yasser Regragui, Lahsen Abouenour, Fettoum Krieche, Karim Bouzoubaa, and Paolo Rosso. 2016. Arabic wordnet: New content and new applications. In *Proceedings of the Eighth Global WordNet Conference*, pages 330–338.
- Horacio Rodríguez, David Farwell, Javi Farreres, Manuel Bertran, Musa Alkhalifa, M Antonia Martí, William Black, Sabri Elkateb, James Kirk, Adam Pease, et al. 2008. Arabic wordnet: Current state and future extensions. In *Proceedings of The Fourth Global WordNet Conference, Szeged, Hungary*, pages 387–405.
- Abdulgabar Saif, Mohd Juzaidin Ab Aziz, and Nazlia Omar. 2017. Mapping arabic wordnet synsets to wikipedia articles using monolingual and bilingual features. *Natural Language Engineering*, 23(1):53–91.
- Hee-Cheol Seo, Hoojung Chung, Hae-Chang Rim, Sung Hyon Myaeng, and Soo-Hong Kim. 2004. Un-supervised word sense disambiguation using wordnet relatives. *Computer Speech & Language*, 18(3):253–273.
- Advait Siddharthan, Nicolas Cherbuin, Paul J Eslinger, Kasia Kozłowska, Nora A Murphy, and Leroy Lowe. 2018. Wordnet-feelings: A linguistic categorisation of human feelings. *arXiv preprint arXiv:1811.02435*.
- Krešimir Šojat, Matea Srebačić, and Marko Tadić. 2012. Derivational and semantic relations of croatian verbs. *Journal of Language Modelling*, pages 111–142.
- Nasrin Taghizadeh and Hesham Faili. 2016. Automatic wordnet development for low-resource languages using cross-lingual word. *Journal of Artificial Intelligence Research*, 56:61–87.
- Piek Vossen. 1998. A multilingual database with lexical semantic networks. *Dordrecht: Kluwer Academic Publishers*. doi, 10:978–94.
- A Zouaghi, L Merhbene, and M Zrigui. 2011. Word sense disambiguation for arabic language using the variants of the lesk algorithm. *WORLDCOMP*, 11:561–567.
- Anis Zouaghi, Laroussi Merhbene, and Mounir Zrigui. 2012. Combination of information retrieval methods with lesk algorithm for arabic word sense disambiguation. *Artificial Intelligence Review*, 38(4):257–269.