Survey and Classification of Methods for Building a Semantic Annotation

Georges Lebbos*, Abd El Salam Al Hajjar*, Gilles Bernard†, Mohammad Hajjar*

*Lebanese University, University Institute of Technology
Saida, Lebanon
†LIASD, University of Paris 8
Saint-Denis, France

Abstract—Though Arabic is one of the five most spoken languages, little work has been done on building Arabic semantic resources. Currently, there is no agreed-upon method for building such a reliable Arabic semantic resource. The purpose of this paper is to present a comprehensive survey of different methods for building or enriching Arabic semantic resources; to study and analyze each method; and to categorize the methods according to their properties. This work should contribute to the definition of new methods and help researchers on Arabic semantics to fit their work in the panel of existing ones.

Keywords—*Lexical semantics; WordNet; Arabic WordNet; Arabic corpus; synset; Arabic semantic resources; translation-based methods; ontologies*

I. INTRODUCTION

Arabic is the first language of more than 200 million individuals across the world and the fifth most spoken one. Surprisingly little work has been done on the building of Arabic semantic resources. Semantics is one of the major components in natural language processing and plays a very important role in improving the performance of information retrieval systems. So there is good motivation to work on such resources.

Lexical semantic relations are associations between meanings of words. For example the nouns "school" (when denoting a building) and "schoolhouse" have a synonymic relation [1]. Semantic resources are lexical databases containing words and semantic associations between them. Arabic semantic resources are scarce; one can cite Arabic WordNet [2] or improved Structured and Progressive Electronic Dictionary for the Arabic language (iSPEDAL) [3].

The aim of this paper is to survey and classify a comprehensive set of methods for building and/or enriching Arabic semantic resources. Each method uses its own methodology and algorithm, usually on relational databases. We show that all existing methods are founded on five types of resources: dictionaries, translation resources, WordNets, ontologies and morphological information, and that this fact makes way for new methods on Arabic corpora resources.

This paper is organized as follows. Section II will review the basic resources used in building Arabic semantic resources, Section III will present a survey of the existing methods and give a detailed description of the procedure for each method. The last section concludes with the classifications of these methods according to their resources and a discussion.

II. MAIN BASIC RESOURCES

This section will describe the main existing accessories, funds and assets used in building most Arabic semantic resources methods, such as WordNets, ontologies, Arabic dictionaries and translation, including morphological resources.

A. WordNet

WordNet is a lexical database representing concepts and their relations, which was developed by linguists and psychologists in 1985 at the Laboratory of Cognitive Science of the University of Princeton [4]. It provides a large repository of English lexical items, and is available online. WordNet was designed to establish relations between the four main parts of speech (POS): noun, verb, adjective and adverb [5]. As similar databases have been devised for other languages (German, French and other European languages, Arabic, among others) "WordNet" is now a generic noun and the original WordNet is here called Princeton WordNet (PWN).

The atomic component of any WordNet is the concept or synset. It is represented as a list of synonymous word forms denoting one sense or a particular purpose; a word can appear in different synsets in case of polysemy, as in the following examples:

- synset$_1$: {car, auto, automobile, machine, motorcar},

- synset$_2$: {car, railcar, railway car, railroad car}.

Synsets are related to each other by semantic relations, such as antonymy, hyponomy, and meronymy.

1) BalkanNet: BalkanNet [6] is a multilingual lexical database relating individual WordNets for the Balkanic languages (Bulgarian, Greek, Romanian, Turkish and Serbian), with the objective of creating a large-scale linguistic resource. It implemented for the first time a new feature, the Inter-Lingual-Index (ILI), to ensure linking of conceptual equivalencies across WordNets in the development of an inter-networked WordNet Management so that each partner retains full responsibility and independence of his local WordNet while simultaneously being able to view other WordNets and check their compatibility.

2) EuroWordNet: EuroWordNet [7] is a multilingual database with individual WordNets for several European languages (Dutch, Italian, Spanish, German, French, Czech and Estonian). Each WordNet is structured in the same way as PWN, with synsets and their relations. It represents a unique

language-internal system of lexicalizations. In addition, individual WordNets are linked by the Inter-Lingual-Index to PWN. Via this index, languages are interconnected so it is possible to go from words in one language to similar words in any other language. The index also gives access to a shared top-ontology (Section II-B) with 63 semantic distinctions. This ontology provides a common semantic framework for all languages, while language specific properties are maintained in individual WordNets.

3) Arabic WordNet (AWN): Arabic WordNet is a lexical database that was created manually in 2006 by [2], [8]. It is one of the few open Arabic semantic resources, and it is based on the design and contents of PWN. Like EuroWordNet, Arabic WordNet is linked to PWN by ILI with the help of the SUMO ontology (see Section II-B).

The first version of AWN contained 9,698 synsets with 21,813 words. A new version has been made available in 2008 [9], which contains 11,269 synsets, with 23,481 words and new named entities.

In 2015, to improve the coverage and usability of AWN, [10] defined its extended version, AWN:V2. This version can be downloaded from the Open Multilingual Wordnet [1]. It includes 37,342 lemmas, 2,650 irregular plurals and 14,683 roots. It also includes 10,169 synsets of which 253 are non-diacritized. The database scheme has been changed and the data cleaned.

AWN:V2 is presented in two formats: Lexical Markup Framework and Lemon Format. The first format is used for standardizing lexical representations of language resources, while the second one is used for linking lexical resources and ontologies.

Despite the efforts made to improve AWN, it remains limited compared to other WordNets.

B. Ontology

There has been, as early as in ancient Egypt, philosophical works describing all beings in the universe and distributing them into categories, as in Porphyry's famous work. Ontology is used in philosophy to designate the study of beings and the nature of being. So, when artificial intelligence needed to coin a word to designate the formal descriptions of beings and relations they created, "ontology" was adopted. For a formal definition, one refers usually to [11], clarified by [12] and [13].

Due to the difficulty of agreeing on a general ontology, ontologies are usually restricted to specific domains. Whether WordNet is such a general ontology is a hot debate; whatever the case, the global architecture is very similar: an ontology is a set of concepts linked by semantic relations; the difference lies in their respective uses.

The Suggested Upper Merged Ontology (SUMO), owned by IEEE, is described in its portal[2] as "the largest publicly available formal ontology today [...] written in SUO-KIF language". SUO-KIF is a simplified derivative of Knowledge Interchange Format, a language equivalent to first-order logic. A translation to OWL (the semantic web language) is also

available. The set has 20,000 terms and 60,000 axioms. There is a complete mapping (by hand) of SUMO to the different versions of PWN (up to version 2.1), including MILO (middle level ontologies) [14, p.73].

This mapping of SUMO concepts to all of the PWN synsets [15] has been very useful; SUMO concepts were also mapped by hand to AWN synsets; thus SUMO plays the role of intermediary between these two WordNets [2], [8]. SUMO is used to maximize the semantic consistency of hyperonymy links. New formal terms will be defined in order to cover a greater number of equivalence associations. These definitions of new terms, in turn, are dependent on the existence of fundamental concepts in SUMO.

C. Arabic Dictionary

According to Oxford Living Dictionary, a dictionary is "a book or electronic resource that lists the words of a language (typically in alphabetical order) and gives their meaning, or gives the equivalent words in a different language". The general structure of dictionaries can be written as {Key = Description}, where the key is a word in some language and description is a set of words which defines, explains and/or provides information such as the key synonyms and antonyms. Available Arabic dictionaries are of two kinds.

- Classical Arabic dictionary
  A classical Arabic dictionary is a paper-based monolingual dictionary. The keys are words in Arabic and the descriptions are also in Arabic. Arabic has a very rich variety of monolingual dictionaries, such as the famous Lisan-Al-Arab or Muajam Makayse Al Luga [16].

- Structured electronic Arabic dictionary
  A structured electronic Arabic dictionary is a monolingual resource. This kind of dictionary contains Arabic words ordered according to a specific logic. Two notable examples:
  - iSPEDAL (improved Structured and Progressive Electronic Dictionary for the Arabic Language) relational database [3] has been designed to be easy to use, with its appropriate query language. For each word it provides links to root, affixes and possible models (patterns). Morphological derivation has much more influence on semantics in Arabic than for instance in English. In addition, there are semantic links between words (definition, spelling, meaning, synonyms, antonyms, usage patterns...). iSPEDAL has been produced from digitized monolingual paper dictionaries, and contains a large number of unused words that need to be verified.
  - LOGOS[3] is a relational database of Arabic verbs with 944 fully conjugated ones.

- Multilingual translation resources
  There are quite a number of bilingual translation resources, from standard paper dictionaries, to New Mexico State University (NMSU) and Al-Mawrid

---

[1]http://compling.hss.ntu.edu.sg/omw
[2]http://www.adampease.org/OP/.

[3]http://www.logosconjugator.org.

Arabic-English databases [17]. Aligned translation corpora are of course of importance here. Though not truly a translation resource, Wikipedia, a project of universal encyclopedia, is a multilingual resource with links that join equivalent resources in different languages.

III. METHODS FOR BUILDING AN ARABIC SEMANTIC RESOURCE

A. Semi-automatic extension of AWN by translation

In 2008 the authors of AWN proposed a method for extending it by bidirectional transaction of part of speech (POS) [18]. They built lists of <English word, Arabic word, POS> tuples from several publicly available English/Arabic translation resources. After cleaning and standardizing the entries, they merged all the lists (using both directions of translation from English to Arabic and vice versa) into one single bilingual lexicon that contained an Arabic or English word and its translation. They then took the intersection of this lexicon with the set of base concept word forms obtained by merging base concepts of EuroWordNet (1024 synsets) and of Balkanet (8516 synsets). Keeping only the tuples whose English word was included in the merged set, they produced <English word, Arabic word, Concept> tuples. Arabic words linked to the same concept were candidates to enter a synset in AWN.

The candidates were to be validated by lexicographers using a methodology with eight heuristic procedures that had been devised while building Spanish WordNet (part of EuroWordNet). This methodology assigned a score to each association between a word and a PWN synset based on some Arabic English bilingual lexicon; AWN being hand-made, no threshold was set and all associations were provided to the lexicographer for verification.

Thus, when editing an Arabic synset, the lexicographer begins with a suggested association, rather than an empty synset with only the English data to go by. Some suggestions were correct or very similar to correct ones. Others were incorrect but served to trigger an Arabic word that might otherwise have been missed. The result has been a much richer set of Arabic synsets.

Initially 15,115 associations were suggested, of which only 9748 (64.5%) have been thus far checked by the lexicographers. The results show that of these, 392 candidates (4.0%) were accepted without any changes, 1246 (12.8%) were accepted with minor changes (such as adding diacritics), 877 (9.0%), while good candidates, were rejected because they were identical or very similar to associations that had already been chosen by the lexicographer, and 7233 (74.2%) were rejected because they were incorrect given the gloss and examples.

B. Semi-automatic Extension of Arabic WordNet using Lexical and Morphological Rules

The same team obtained better results with another method [9] which derived all possible forms from Arabic words in AWN (thus producing many non-existent words), tested them with existing databases, looked for their translation in English, linked them to PWN and then back to AWN, to be manually

validated as in Section III-A. More precisely, their procedure can be described into two phases. The first phase includes:

- Collect a set of validated basic verb forms from AWN (in Arabic, most noun and adjective forms are derived from verb forms).

- Apply to this set all Arabic derivational morphological patterns and generate all possible stems.

- Attach affixes to each produced stem according to Arabic morphological rules and generate all possible words.

- Control the existence of each word with GigaWord non free Arabic corpus, the LOGOS multilingual translation portal, and the NMSU Arabic-English lexicon.

- Associate each attested word to its English translations in the last two portals (Logos and NMSU).

In the second phase, they had to devise a method for linking each word to an AWN synset. Each word had various translations in English, each English word was potentially linked with more than one PWN synsets, and PWN synsets can be more or less closely related by semantic links. Thus, in order to automatically assess if some Arabic words were semantically related, those three types of relations had to be considered. This was done with the following procedure:

- Collect the set of <Arabic word, English word, PWN synset> tuples for a given Arabic base verb form and its derivatives.

- Extract the set of English synsets and identify all existing semantic relations between these synsets in PWN.

- Build a graph with three levels of nodes corresponding to Arabic words, English words, and English synsets, respectively and edges corresponding to the translation relationships between Arabic words and English words, the membership relations between English words and Princeton WordNet synsets and finally, the relations between Princeton WordNet synsets.

Finally, they manually review the candidates and include the valid associations in AWN.

They randomly selected 10 of the 2296 verbs currently in AWN to manually control the results.

C. Extending Named Entities Coverage of AWN using Wikipedia

Other members of the same team extracted named entities from Arabic Wikipedia, linked them to named entities from the corresponding English Wikipedia page, linked those to named entities from PWN, and then back to synsets of AWN [19]. Though the result was much better, the coverage was scarce. This was done using the following procedure:

- Extract some English Named Entity from Princeton WordNet.

- Generate a couple <English NE, Arabic NE> using the interwiki link from English Wikipedia to Arabic Wikipedia.

- Manually vocalize Arabic NE.

Experiments showed that 93.3% of NE synsets were correct. But the size of the automatically evaluated set was small (only 496 synsets, or 12% of the set of the recovered synsets). From the 3,854 proposed assignments, 3,596 (93.3%) were correct, 67 (1.7%) were wrong and 191 (5%) were not known by the reviewer.

D. Amine Arabic WordNet

A different approach [20], [21] exported AWN into a database integrated with Amine ontology and structured according to the mapping between PWN synsets and SUMO concepts, and added Arabic synonyms.

Amine is an open-source multi-layer Java platform dedicated to the development of intelligent and multi-agents systems. It is a modular environment composed of four layers: i) Ontology layer; ii) Algebraic layer; iii) Programming layer; and iv) Agents and Multi-Agents Systems layer. Amine platform supports intelligent processing with the possibility of defining inference rules, thus giving rise to opportunities for exploring the semantic aspect in automatic processing of languages [22]–[27].

The approach was initiated by the construction of an Amine ontology termed Amine AWN. They exported the entire set of data embedded in AWN tapped by a Java module based on Amine Platform APIs. This module used the mapping between PWN synsets and SUMO concepts to build the Amine AWN type hierarchy. Then, it added Arabic synonyms based on the links between PWN synsets and AWN synsets.

The equivalence is not the only relation which links PWN synsets to the SUMO concepts. In PWN a synset can be a more specialized subset of some general synset, in which case the module creates a new subtype of the SUMO concept. At that moment, AWN synsets are added as synonyms for this new subtype. In the case of "has instance" a new individual is created instead of a subtype. This yields the first level of type hierarchy.

The second level is obtained by a similar processing making use of the relations between PWN synsets. At this stage a hyponymy or hyperonymy relationship is considered as a specialization (or generalization) relation of the previous stage.

In addition, the module allows the automatic extraction of SUMO concepts definitions written in SUO-KIF notation.

E. Arabic WordNet Enrichment by Morphological and Translation

A similar method to that of Section III-B was proposed in 2009 [28]; words were morphologically parsed with rules devised by linguists, then translated and associated to synsets through equivalence relations between the synsets made explicit by the Inter-Lingual Index, which serves as a semantic deep structure [7]. Their model is based on a grammar of templates rules for parsing morphological Arabic data.

The main purpose of the semantic relation module is to provide a common framework for the most important concepts shared between a ll the WordNets. It consists of basic semantic distinctions that classify a subset of ILI representing the most important concepts in the related WordNets.

The methodology is based on the integration of other Word-Nets with AWN. The proposed model experiments this process on sequences of string matching and string manipulation steps. The morphological parsing templates are made by a domain expert in order to find root(s) and associated features.

F. Enrichment of Arabic WordNet using YAGO Ontology

The same team that elaborated Amine AWN (section III-D) used YAGO ontology from Max-Planck Institute, translating its named entities into Arabic with Google translation, then added them to AWN according to two types of mappings (direct mapping through WordNet, mapping through YAGO relations to AWN synsets) [27], [29]. Their use of YAGO is justified by the fact that it contains named entities (NE) already identified and checked.

YAGO (Yet Another Great Ontology) is a large ontology with high coverage and precision with the following features:

- It covers a great amount of individuals (2 million NEs).

- It has a near-human accuracy, around 95%.

- It is built from WordNet and Wikipedia.

- It is connected to the SUMO ontology.

- It exists in many formats (XML, SQL, RDF, etc.).

- It is available with tools that facilitate exporting and query.

The first step translates all YAGO entities from English into Arabic. This translation is performed automatically using Google Translation API (GTA). The translated YAGO entities have been added in Arabic WordNet according to two types of mappings as follows [27]:

- Direct mapping through PWN
  The PWN synsets corresponding to a YAGO entity are identified using the TYPE relation in the YAGO facts. After that, the AWN synsets corresponding to the identified PWN synsets are connected with this entity. "For example, the YAGO entity "Abraham Lincoln" appears in three facts for the "TYPE" relation; from these facts, the three English WN synsets "president", "lawyer" and "person" are extracted. Hence, the YAGO entity ابراهام لنكولن (i.e., Abraham Lincoln) can be added as an instance corresponding respectively to AWN synsets identified by رئيس (president), شخص، انسان (person, human) and محام، محامي، وكيل (lawyer, attorney)" [27].

- Mapping YAGO relation / Arabic WordNet synsets: A mapping is performed between arguments of YAGO relations and instances of AWN synsets; certain types of argument were previously manually linked to specific synsets.

They conclude that "[a]fter applying this technique to the three million YAGO entities, we found it was possible to keep 433,339 instances (145,135 NEs thanks to the first mapping and 288,204 NEs from the second mapping) that were connected with 2,366 corresponding AWN synsets. This number represents around 38,000 times the number of existing NE instances in AWN." [27]

G. Enriching AWN from Aligned Multilingual Corpus

Abdul Hay's PhD thesis [30] extracted semantic categories from a multilingual aligned corpus with English and languages from EuroWordNet. If all but Arabic words were members of synsets linked by Inter-Lingual Index, then the Arabic word should also be in a linked synset in AWN.

This thesis had a more general objective: implement and evaluate techniques for extracting semantic relations from a multilingual aligned corpus, enrichment of AWN being one - important - outcome.

The corpus is in four languages: French, Arabic, English and Spanish. The proposed method begins with a phrasal alignment of the corpus and then extracts translation equivalences. It uses the idea of finding "cliques" in this aligned multilingual corpus in order to extract concepts or synsets and input them to WordNet.

Cliques are maximally connected sub-graphs where all units are interconnected due to possible semantic intersections. They have the advantage of giving information on both the synonymy and polysemy of units, and providing a form of semantic disambiguation. An example of clique might be the set of nouns {Ar. قسم, Fr. fragment, En. snippet, Sp. recorte}. Cliques can be connected with EuroWordNet in order to evaluate the possibility of the recuperation of the semantic relation for the Arabic units already declared in the English, French and Spanish units that exists in the Inter-Lingual Index (ILI). If all the units (English, French, Spanish) share one sense in EuroWordNet (via ILI) then the Arabic units are one synset that will be inserted in Arabic WordNet. Based on the thesis results, 84% of the extracted synsets are accurate and measured manually.

H. Semantic Enrichment of iSPEDAL based on Arabic Dictionary

iSPEDAL, proposed in 2010 [31], [32], includes an automatic system that enriches it with morphological information from classic dictionaries or from any Arabic textual corpus. In 2013 a heuristic has been proposed by the iSPEDAL team for automatically enriching it with semantic information [3], using semi-structured information from plain standard dictionaries to deduce semantic links (synonymy, antonymy).

It begins with searching for traditional keywords that are usually used to introduce some sort of semantic commentary on the word or root in question, then propose hand-made rules for deducing relations. The following examples are taken from the Lisan-Al-Arab dictionary.

- The word أي (meaning) is found under the root أكل (food) preceeding a synonym of the root, طعام (food),

so can be considered as a keyword. Under the root كتب (to write) , the noun كاتب (writer) is followed by the word جمعها (plural), followed in its turn by the plural form كتاب (writers). The word جمعها can be considered as a keyword; the same goes for عكسها (opposite) that can be considered as a keyword for antonym.

- Colons can be considered as keywords, but their use has to be manually desambiguized. A colon can link root to derivatives, as the colon separating the root عرف (know) from the expression عرف الضالة (recognition of the ignorant). Or it can link two words semantically related. Under the same root عرف (know), a colon separates the derived word العرفان (recognition) and العلم (science); considered as near synonyms.

I. Semantic Enrichment of iSPEDAL based on Translation

The same team proposed another method for enriching iSPEDAL using translation by available resources to and from a foreign language to compute synonymy of Arabic words by correlating their translations [3]. In practice they used English - Arabic and Arabic - English translation resources. If two Arabic words have the same sense, probably they have the same translation in English. The authors begin by translating two Arabic words in English, yielding for each word a set of English words. Then they calculate a similarity factor between both words by computing the number of common words divided by the total number of words in the sets. This factor is then used as a threshold to consider the two words as synonyms, no result has been given for this method.

J. Extracting Semantic Relations from Arabic Wiktionary

Another approach [33] extracts synonymy and antonymy relationships from Arabic Wiktionary. The procedure is as follows:

- Preprocessing phase with extraction of definitions.

- Analysis of the vocabularies of these definitions and extraction of the semantic relations induced. They used the segmentation tool AraSeg [34] to cut the texts of the definitions into lexical units. Then a morphosyntactic analyzer uses the different knowledge resources of the Arabic language to extract the lemmas and the grammatical classes of the words in the definitions.

- Creating a lexical database, linking the words and the semantic knowledge found in the previous phase in order to construct the general structure of the data [33].

K. Arabase

Arabase platform [35] aims to integrate every available (in 2013) Arabic semantic resource, from *King Abdulaziz City for Science and Technology* (KACST) database, to Arabic StopWords Sourceforge resource[4] and AWN. It has, according

---

[4]This resource is a simple hand-made list of stopwords that we did not consider as needing a section by itself.

to the authors, "a good potential to interface with WordNet". Arabase computes by hand-made rules semantic properties of vocalized words[5] and forms a sort of virtual WordNet, so the final vector representing a word could be:

$$<I_m, I_{se}, I_{sy}, I_{sm}, I_{WN}>$$

Where, for a given word,

- I_m is the morphological analysis: its root, category and genus.

- I_{se} contains its meanings with definitions for each one.

- I_{sy} designates its relations with sets of words, as synonyms, antonyms ....

- I_{sm} designates the semantic fields the word belongs to, with the semantic relations between the semantic fields.

- I_{WN} indicates the English word equivalents in PWN the word is related to.

The process of integration takes place in four stages [35]:

1) Analysis: Unify the format of all resources and transfer them into a single MYSQL database.
2) Design an integrated target database: The scalable design of KACST database was the starting point of the integrated database, they changed it afterwards to match.
3) Integration: Apply an algorithm to automatically compile these resources together. The main input of the embedded resources is the unvocalized word that has more than one vocalized form. These vocalized forms can be names, verbs, particles or unclassified.
4) Linking: Group all unvocalized words by meaning.

IV. CONCLUSIONS

We have analyzed here a comprehensive list of published methods for building Arabic semantic resources. Table I introduces a characterization of these methods according to the resources used by each one.

TABLE I.   METHODS FOR BUILDING ARABIC SEMANTIC RESOURCES

Section	Dictionary	Translation	WordNets	Morphology	Ontology
III-A		X	X		
III-B		X	X		
III-C	X	X	X	X	
III-D		X	X		X
III-E	X	X	X	X	
III-F		X	X		X
III-G		X	X		
III-H	X				
III-I		X			
III-J	X			X	
III-K	X	X	X	X	X

This table makes obvious that the translation resources are the most used ones, closely followed by WordNets, mostly PWN but also others. At the other end, ontologies appear to be the less used resources. On another side, one can see that some methods are concentrated on only one resource (as III-H or III-I), while III-K for instance combines a large set

of resources. Various classifications of these methods can be produced according to the resources used.

But the most interesting fact is what this table does not contain. It shows that researches on Arabic semantic resources have extensively used foreign resources (translation, WordNets and ontologies), but very little has been done on extracting semantic information from Arabic data alone (only method III-J), and nothing on extracting information from large Arabic corpora, though these are now available.

This is probably where lies the next step for new methods: build or enrich Arabic semantic resources with probabilistic procedures taking as input large Arabic corpora. As a first element in that direction, we contributed with such a procedure to enrich AWN in [36].

REFERENCES

[1] G. Lebboss, M. Hajjar, and G. Bernard, "Wordnet and arabic lexical semantics," in *IARIAS 2014 International Multidisciplinary Conferences on Industrial and Applied Sciences*, S. B. Sadkhan, Ed., vol. 1, University of Babylon.   Dubai, UAE: IARIAS, apr 2014, pp. 48–61.

[2] S. Elkateb, W. Black, H. Rodrguez, M. Alkhalifa, P. Vossen, A. Pease, and C. Fellbaum, "Building a wordnet for arabic," in *In proceedings of the Fifth International Conference on language Ressources and Evaluation*.   Genoa,Italy: Citeseer, 2006.

[3] M. Hajjar, A. E. S. Al Hajjar, Z. Abdel Nabi, and G. Lebboss, "Semantic enrichment of the iSPEDAL corpus," in *3rd World Conference on Innovation and Computer Science (INSODE)*, 2013.

[4] G. A. Miller, "Wordnet: A dictionary browser," in *Information in Data, Proceedings of the First Conference of the UW Centre for the New Oxford Dictionary, Waterloo Univ.*, 1985.

[5] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller, "Introduction to wordnet: An on-line lexical database," *International journal of lexicography*, vol. 3, no. 4, pp. 235–244, 1990.

[6] S. Stamou, K. Oflazer, K. Pala, D. Christoudoulakis, D. Cristea, D. Tufis, S. Koeva, G. Totkov, D. Dutoit, and M. Grigoriadou, "Balkanet: A multilingual semantic network for the balkan languages," in *Proceedings of the International Wordnet Conference, Mysore, India*, 2002, pp. 21–25.

[7] P. Vossen, "EuroWordNet: a multilingual database of autonomous and language-specific WordNets connected via an Inter-Lingual Index," *International Journal of Lexicography*, vol. 17, no. 2, pp. 161–173, 2004.

[8] S. Elkateb, W. Black, P. Vossen, D. Farwell, A. Pease, and C. Fellbaum, "Arabic wordnet and the challenges of arabic," in *Proceedings of Arabic NLP/MT conference, London, UK*.   Citeseer, 2006.

[9] H. Rodriguez, D. Farwell, J. Farreres, M. Bertran, M. Alkhalifa, M. A. Martí, W. Black, S. Elkateb, J. Kirk, A. Pease, P. Vossen, and C. Fellbaum, "Arabic WordNet: Current state and future extensions," in *Proceedings of The Fourth Global WordNet Conference, Szeged, Hungary*, no. 387–405, Marrakech (Morocco), 2008.

[10] Y. Regragui, L. Abouenour, F. Krieche, K. Bouzoubaa, and P. Rosso, "Arabic wordnet: New content and new applications," in *Proceedings of the Eighth Global WordNet Conference*, Bucharest, Romania, January 27-30 2016, pp. 330–338.

[11] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing?" *International Journal of Human-Computer Studies*, vol. 43(5-6), pp. 907–928, 1995.

[12] D. L. McGuinness, R. Fikes, J. Rice, and S. Wilder, "An environment for merging and testing large ontologies," in *KR*, 2000, pp. 483–493.

[13] C. Feilmayr and W. W, "An analysis of ontologies and their success factors for application to business," *Data & Knowledge Engineering*, 2015.

[14] F.-R. Chaumartin, "Wordnet et son cosystme : un ensemble de ressources linguistiques de large couverture," in *Colloque BD lexicales*, Montréal, Canada, 2007, p. 59. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00611240

---

[5]Arabic words are written without short vowels in normal use and in a majority of documents.

[15] I. Niles and A. Pease, "Linking lexicons and ontologies: Mapping Word-Net to the Suggested Upper Merged Ontology," in *Proceedings of the International Conference on Information and Knowledge Engineering (IKE 03), Las Vegas, Nevada*, vol. 2, Las Vegas, Nevada, USA, 2003, pp. 412–416.

[16] L. Sakher, *Lisan Al-Arab, Al Qamous Al Mouhit, Al Wasit, Al Mouhit, Mouhit Al Mouhit, Al Ghani, Taj Al Arous, Najaat Al Raed*, L. Sakher, Ed.   http://lexicons.sakhr.com: Sakher, Lexicons, 2009.

[17] R. Baalbaki, *Al-Mawrid Arabic-English*.   Dar al-Ilm Lilmalayin, 2012.

[18] H. Rodrguez, D. Farwell, J. Farreres, M. Bertran, M. Alkhalifa, and A. Mart, "Arabic wordnet semi-automatic extensions using bayesian inference," in *LREC*, Marrakech (Morocco), 2008.

[19] M. Alkhalifa and H. Rodriguez, "Automatically extending named entities coverage of Arabic WordNet using Wikipedia," *International Journal on Information and Communication Technologies*, vol. 1, no. 1, pp. 1–17, June 2008.

[20] L. Abouenour, K. Bouzoubaa, and P. Rosso, "Improving Q/A using Arabic WordNet," in *Proc. of the 2008 International Arab Conference on Information Technology (ACIT'2008), Tunisia*, December 2008.

[21] ——, "Système de question/réponse dans le cadre d'une plateforme intégrée: cas de l'arabe," in *Proc. Rencontre Nationale en Informatique: Outils et Applications, RINOA-2008, Errachidia, Morocco*, jun 2008, pp. 5–7.

[22] A. Kabbaj, K. Bouzouba, and A. Soudi, "Amine platform: an artificial intelligence environment for the development of intelligent systems," *Artificial Intelligence*, 2005.

[23] A. Kabbaj, K. Bouzouba, K. El Hachimi, and N. Ourdani, "Ontologies in amine platform: structures and processes," in *Conceptual Structures: Inspiration and Application*.   Springer, 2006, pp. 300–313.

[24] A. Kabbaj and K. Bouzouba, "Amine: une plate-forme pour le développement de systèmes et dagents intelligents," *e-TI , Revue lectronique des Technologies de l'Information*, vol. 1, oct 2005. [Online]. Available: http://www.revue-eti.net/document.php?id=364

[25] A. Kabbaj, "Development of intelligent systems and multi-agents systems with amine platform," in *Conceptual Structures: Inspiration and Application*.   Springer, 2006, pp. 286–299.

[26] L. Abouenour and K. Bouzoubaa, "Using an arabic ontology to improve the q/a task," *the 11th International Business Information Management Association Conference IBIMA,*, 2009.

[27] L. Abouenour, K. Bouzoubaa, and P. Rosso, "On the evaluation and improvement of Arabic WordNet coverage and usability," *Lang Resources & Evaluation*, vol. 47, no. 3, pp. 891–917, jun 2013. [Online]. Available: http://dx.doi.org/10.1007/s10579-013-9237-0

[28] H. M. Al-Barhamtoshy and W. H. Al-Jideebi, "Designing and implementing arabic wordnet semantic-based," in *the 9th Conference on Language Engineering*, 2009, pp. 23–24.

[29] L. Abouenour, K. Bouzoubaa, and P. Rosso, "Using the Yago ontology as a resource for the enrichment of named entities in Arabic WordNet," in *Proceedings of The 7th International Conference on Language Resources and Evaluation (LREC 2010) Workshop on Language Resources and Human Language Technology for Semitic Languages*, 2010, pp. 27–31.

[30] A. Abdulhay, "Constitution d'une ressource sémantique arabe à partir d'un corpus multilingue aligné," Ph.D. dissertation, Université de Grenoble, 2012.

[31] A. E. S. Al Hajjar, "Extraction et gestion de l'information à partir des documents arabes," Ph.D. dissertation, Paris 8 University, 2010.

[32] M. Hajjar, A. e. S. Al Hajjar, K. Zreik, and P. Gallinari, "An improved structured and progressive electronic dictionary for the arabic language: iSPEDAL," in *Proc. of the fifth International Conference on Internet and Web Applications and Services*, 2010.

[33] B. Abdelali and Y. Tlili-Guiassa, "Extraction des relations sémantiques à partir du Wiktionnaire arabe," *Revue RIST*, vol. 20, no. 2, pp. 47–56, 2013.

[34] Z. Mouelhi, "Araseg : un segmenteur semi-automatique des textes arabes," *In JADT 2008 : 9es Journes internationales dAnalyse statistique des Donnes Textuelles*, vol. 1, pp. 867–878, 2008.

[35] H. Raafat, M. Zahran, and M. Rashwan, "Arabase: A database combining different arabic resources with lexical and semantic information," in *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval and the International Conference on Knowledge Management and Information Sharing*.   Scitepress, 2013. [Online]. Available: http://dx.doi.org/10.5220/0004656702330240

[36] G. Lebboss, G. Bernard, N. Aliane, and M. Hajjar, "Towards the enrichment of arabic wordnet with big corpora," in *Proceedings of the 9th International Joint Conference on Computational Intelligence*, vol. 1, November 2017.