Derivational Relations in Arabic WordNet

Mohamed Ali Batita
Research Laboratory in Technologies of Information and Communication & Electrical Engineering,
Tunisia
BatitaMohamedAli@gmail.com

Mounir Zrigui
Research Laboratory in Technologies of Information and Communication & Electrical Engineering,
Tunisia
Mounir.Zrigui@fsm.rnu.tn

Abstract

When derivational relations deficiency exists in a wordnet, such as the Arabic WordNet, it makes it very difficult to exploit in the natural language processing community. Such deficiency is raised when many wordnets follow the same development path of Princeton WordNet. A rule-based approach for Arabic derivational relations is proposed in this paper to deal with this deficiency. The proposed approach is explained step by step. It involves the gathering of lexical entries that share the same root, into a bag of words. Rules are then used to affect the appropriate derivational relations, i.e. to relate existing words in the AWN, involving part-of-speech switch. The method is implemented using Java. Manual verification by a lexicographer takes place to ensure good results. The described approach gave good results. It could be useful for other morphologically complex languages as well.

1 Introduction

A wordnet is a lexical database built of synsets. One synset represents one concept and contains words from the same part of speech (POS) (noun, adjective, verb, and adverb). Synsets are interconnected with different relations. But, there are no cross-part-of-speech relations. This type of relation is a link between words sharing the same stem and meaning like the verb 'eat' and the noun 'eater'. The first WordNet, Princeton WordNet (Fellbaum, 2010), was built for the English language. Since that, many wordnets has seen the light for over 160 languages[1]. One of them is the Arabic WordNet (Elkateb et al., 2006) (henceforth, AWN) for the Modern Standard Arabic. AWN followed the development of Princeton WordNet and EuroWordNet (Vossen, 1998).

Started in 2007, researches on AWN are made to improve it. Some of theme improved its contents (Boudabous et al., 2013; Saif et al., 2015). Others used it in different disciplines of the Natural Language Processing (henceforth, NLP) (Abouenour et al., 2008; Abouenour et al., 2013). Despite the greatness of these works, it clearly did not take into consideration the specificities of the Arabic language, especially, its morphological aspect.

A lexicon, like AWN, needs to have an extensive coverage, high quality, and multiple use in NLP applications (Mallat et al., 2015a) (Mallat et al., 2015b) (Ayadi et al., 2014) (Mohamed et al., 2015). Adding to that, derivational morphology provides handful information for the benefit of the NLP. As proof, Wilbur and Smith(Wilbur and Smith, 2013) showed that it can be used to calculate probabilities of semantic relatedness. Also, Sagot (Sagot, 2010) used derivational analysis to determine if an unknown word can be used to create a new one for a lexicon extension. Derivational morphology is used to extend different wordnets like Bulgarian, Serbian and Romanian WordNet(Koeva et al., 2008; Koeva, 2008; Mititelu, 2012). The aim of this pilot study is to enrich the AWN with derivational relations using a rule-based approach to extend its coverage and turns it into a more useful knowledge base.

The rule-based approach includes domain knowledge into linguistic knowledge. This yield accurate results. Yet, linguistic knowledge acquired for one NLP system may be reused to build others systems that require similar knowledge. Those approaches are based on a solid core of linguistic knowledge. They depend on hand-constructed rules from a lexicographic rather than automatically gathered from data.

[1]Extended Open Multilingual WordNet: http://compling.hss.ntu.edu.sg/omw/summx.html

This paper is structured into five sections. Section 2 is an overview of the AWN. Section 3 will provide some background on the Arabic language. In section 4, we will discuss some related works. We will also discuss the choice of the rule-based approach regarding other approaches. We will speak about our approach in details in section 5. Last but not least, we will show the obtained results in section 6.

2 Overview of the Arabic WordNet

AWN's development followed the top-down procedure. It consists of translation the Princeton WordNet's core and extending it through more specific concepts related to the Arabic culture. This procedure expands the compatibility between wordnets. It is based on manually encoding of the specific concepts. The first version V1 of AWN contains 21,813 words grouped into 9,698 synsets, 6 types of relations between those synsets corresponding to 143,715 links. The second version V2 released in 2008 containing new synsets and links. It contains 11,269 synsets, equivalent to 23,481 words, and 161,705 links equivalent to 22 types (5 of them for the interconnection between PWN and AWN) (Batita and Zrigui, 2017). Another version is freely available on the internet structured under the Lexical Markup Framework (LMF) developed by (Abouenour et al., 2013). Table 1 below will recapitulate the number of the words, synsets, and links in all the 3 versions of the AWN.

Table 1: AWN's versions.

	V1	V2	LMF
Words	21,813	23,481	60,157
Synsets	9,698	11,269	8,550
Links	143,715 (6 types)	161,705 (22 types)	41,136 (4 types)

We notice that V2 contains more synsets and fewer words then the LMF file. In one hand, V2 has 11,269 synsets related with 161,705 links, on the other hand, the LMF file 8,550 synsets related with only 41,136 links. This is not proportional to the number of the words.

There are many kinds of links in AWN V2. Table 2 displays 17 links. There are 5 others links but they are not between Arabic words. They are inter-language links. We have no interest in them. each one with an example and if it exists in the LMF file or not.

Table 2: Links in AWN.

Link	Example	Frequency	
		V2	LMF
Has hyponym	شراب، ماء *šrāb, mā-ʾ* (drink, water)	9,347	19,806
Has derived	تعليمي، تعليم *tʿlymy, tʿlym* (educational, education)	178	-
Related to	لجأ، ملجأ *lǧʾa, mlǧʾa* (refuge, shelter)	4,769	-
Has holo member	اكل لحم، لواحم *ākl lḥm, lwāḥm* (carnivore, carnivores)	334	-
Near antonym	زيادة، نقصان *zyādt, nqṣān* (increase, decrease)	772	14
See also wn15	وديعة، طلبية *wdyʿt, ṭlbyh* (deposit, order)	166	-
Has holo part	خلية، متعضي *ẖlyt, mtʿḍy* (cell, organism)	697	-
Has holo made of	ورقة، صفحة *wrqt, ṣfḥh* (paper, page)	60	-
Has subevent	وقف، قام *wqf, qām* (stand, stand up)	128	-
Category term	انسان، جسم *ānsān, ǧsm* (human, body)	548	-
Near synonym	أسبق، مبكر *ʾasbq, mbkr* (premier, early)	122	412
Be in state	اتصل، متصل *ātṣl, mtṣl* (contact, connected)	83	-
Has instance	عاصمة، القاهرة *ʿāṣmt, ālqāhrh* (capital, Cairo)	929	549
Verb group	ضرب، صدم *ḍrb, ṣdm* (hit, bump)	142	-
Causes	حرّك، حوّل *ḥrk, ḥwl* (move, displace)	75	-
Region term	بابل، عشتار *bābl, ʿštār* (Babylon, Ishtar)	35	-

Usage term	أسبرين، إسم تجاري *ʔasbryn, ʔism tğāry* (Aspirin, commercial name)	3	-

To clarify, the link *near synonym* is represented in the LMF file by the name *similar*[2] and *near antonym* by just *antonym*. The two links *has hyponym* and *has instance* are splitted into *hyponym/hypernym* and *isInstance/hasInstance* respectively.

If we can classify those links, we can say that there is two types; semantic and derivational link. Semantic links rely on words sharing some meaning. Most of the presented links are semantic like *has holo part, has holo made of, has subevent...* Only two links are morphosemantic links; *has derived* and *related to*. Not only they are morphologically but also semantically relying on words. They rely on words that share the same root but have different POS.

There is a third type of link which is morphosemantic relations. As it is claimed in (Šojat et al., 2012), there is a difference between the derivational and morphosemantic links. The derivational relations are language-specific while the morphosemantic relations are not.

3 Arabic language

As it is widely known, the Arabic language is a Semitic language which makes it different from other languages, like English or French. It is characterized by an inflectional and derivational morphology. Inflectional morphology is divided into verbal and nominal morphology. The verbal morphology bends on the aspect, the mood, the voice and the subject (person, gender, and number) of the verbs. The nominal morphology bends on the gender, the number, the state, and the case of nouns, the adjectives, and the proper nouns. The derivational morphology consists of the deverbal noun, the active participle, the passive participle and other derivations based on patterns change (Habash, 2010). All things considered, this richness provides an effective information for many NLP tasks.

Besides, Arabic is a notable language for its *nonconcatenative* morphology which is the modification of the internal structure of a word. In other words, it is a form of a word in which the root, usually three consonants and called *triliteral root*, is modified by adding other consonants and vowels. Generally, in Arabic, the derivation is based on three concepts. Given a *root* and a *pattern*, you can create a word *form* by applying derivational rules. This makes it difficult to automatically construct new words from a primitive root. For example, the Arabic words دارس *dārs* (student) and مدرس *mdrs* (teacher)[3] share the same Arabic root د ـ ر ـ س *d - r - s* (d - r - s) which is the concept of studying. To that end, we can say that those two words are *derivationally* and *semantically* related. More details about the Arabic morphology, you can found it in (Habash, 2010).

4 Related Work

Even though derivational morphology is a numerous area of studies, we did not found many lexical resources that rely on this kind of morphology, in the Arabic language. Derivational relations enrichment started with the Turkish WordNet in 2004. Bilgin et al. (Bilgin et al., 2004) described a semi-automatic approach to add new synsets by applying derivational rules to existing words and add a morpho-semantic link between them. This type of approach is basically adding automatically suffix and prefix to a steam. Since it is automatic, manually validation is mandatory and important. the same work is done to the Czech WordNet (Pala and Hlaváčková, 2007).

Fellbaum et al. (Fellbaum et al., 2007) did not follow the same approach but instead, he added morphological relations between derived pairs of words in PWN. The derived pairs of words are recognized automatically since they share the same steam. Manual validation is also necessary. This type of relation is cross-POS (between verb and noun pairs). In 2012, the same kind of work is followed in the Romanian WordNet by Mititelu (Mititelu, 2012). The work is summarized in two steps. The first step is to create all possible combination, given 3 lists of words, prefixes, and suffixes. The second step is to validate the affectation of prefixes and suffixes, each one aside, using a set transformation rules.

The Bulgarian (Koeva, 2008), the Serbian (Koeva et al., 2008) and the Polish WordNet (Piasecki

[2]The link *similar* exist in V2 but it is an interlanguage link.

[3]From now on, Arabic words will be followed by their transliteration using the transliteration system of LaTeXand their English translations in brackets.

et al., 2009) adopted another type of approach. Based on the alignment to the PWN, the approach consists of transforming the derivational relations existing in the PWN to each wordnet. In their case of study, Koeva et al. (Koeva et al., 2008) proposed several approaches to make the generating of new synsets and relations possible based on derivational patterns of different POS.

Outside wordnets, *Lefff* (Sagot, 2010) is a morphological lexicon for French based on the lexical framework Alexina. This framework is used with different languages to develop morphological and syntactic NLP lexicons like Persian, Sorani, Kurdish and even English. This lexicon is freely available and has a large coverage. It is constructed by merging several existing resources using semi-automatic techniques and conversion. Remaining with the same language, *VerbAction* (Tanguy and Hathout, 2002) too is a morphological resource who couples verbs with their action nouns (inspect/inspection). *VerbAgent* (Tribout et al., 2012) is the same as *VerbAction* but with agent nouns (inspect/inspector).

The available evidence seems to suggest that the development of those resources is either based on manual work or validation and/or lexical information (derivational and morphological rules). Other attempted researches are less supervised and based only on morphological information. Can et al. (Can and Manandhar, 2009) proposed an unsupervised method based on different POS to produce morphological rules. Bernhard (Bernhard, 2010) described two methods for unsupervised learning of morphological families. The first one is called *MorphoClust*. It clusters words into families using hierarchical classification methods. The second one is called *MorphoNet*. It constructs a lexical network from the words presented in *MorphoClust*. The words represent the nodes and the morphological relations represent the links between those words.

Recently in 2016, Zaghouani et al. (Zaghouani et al., 2016) have developed the AMPN, a semantic resource, based on Arabic morphological patterns. It clusters the verbs of Arabic PropBank[4] (Kipper et al., 2008) according to their morphological patterns. Arabic verbs are studied according to their lemmas. They are defined as a combination of root and morpheme patterns of the verbs.

[4]Annotated corpus with verbal propositions and arguments.

Basically, the cited approaches rely on morphological rules. In another way, they are rule-based approaches. Each one used some morphological rules specific to its language to whether generate new words (adding prefixes and suffixes) or coupling existing words (share the same steam). The advantage of this type of approach is the analysis of the input and output of a system using linguistic knowledge. Also, it helps the language learner's to better understand the language. However, other approaches, like statistical-based or machine learning, cannot distinguish between well-formed or ill-formed input which is an issue in some NLP tasks (Shaalan, 2010).

There is a rapidly growing literature on (Shaalan, 2010), which indicates that rule-based approach for Arabic NLP tasks reported successful results. Shaalan presented 4 tools and 3 systems based on Arabic morphological and syntactic rules. The tools are about morphological analyzer/generator and syntactic analyzer/generator. The 3 systems are Machine Translation, Named Entity Recognition, and Computer-assisted Language Learning. The aim of this study is to show that the development of systems based on rule-based approach is feasible with languages like Arabic (absence of linguistic resources and difficulties of adapting tools from other languages...). All things considered, it seems reasonable to base our work on this kind of approach. Next section will describe precisely each step of our proposed approach.

5 Our Approach

Since there is a lack of derivational relations in AWN, we will attempt to add them based on the existent words in it. The suggested approach depends on lexical entries and some transformation rules. We will gather lexical entries sharing the same root into *bag of words* and we will use the rules to affect the appropriate types of derivational relations. Each rule is based on the POS and the patterns of the words. The following figure 1 shows an Arabic word with its derived forms and each with its pattern (*1, 2, and 3 in the patterns refer to the three consonants of the triliteral root*).

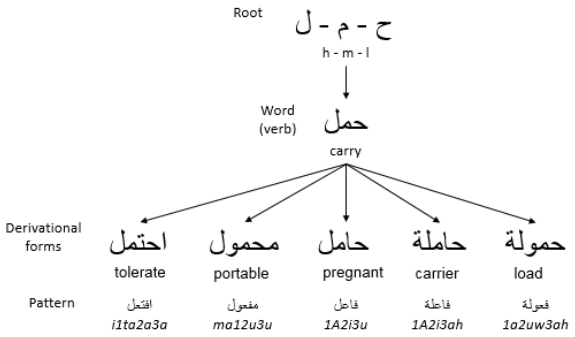

Figure 1: The derivations of the Arabic verb حمل *ḥml* (carry) with their patterns.

The issue under scrutiny in derivational morphology is creating new words from others. In our work, instead of creating new words we will use only words that exist in AWN and try to make a connection between them. This task involves POS switch (sometimes, it is preserved and we will see how). To give an illustration, let's look at the example in table 3. We gain from a verb a noun and from a noun another noun and an adjective.

Table 3: Derivation between part of speeches.

Verb→ Noun	Noun→ Adjective	Noun→ Noun
كتب، كتاب *ktb, ktāb* (write, book)	كتاب، كتابي *ktāb, ktāby* (book, my book)	كتاب، كتيب *ktāb, ktyb* (book, brochure)

One can tell that there is a link between two words if (i) they belong to the AWN (ii) they share the same root and (iii) there is a rule which allows the transformation. Our method is described step by step in the next subsections.

5.1 Clustering Words into Bag of Words

First of all, we gather the words that share the same root in a so-called *bag of words*. This step is based on the root of each word in AWN. This step will help us to:

1. Eliminate the *underivatized* words like named entities ...بنز - مرسيدس، اينشتاين *ā-ynštāyn, mrsyds - bnz...* (Einstein, Mercedes-Benz...) and multiword expression,

2. Keep the *apolistic generic noun* like ...فيل، خروف *ḫrwf, fyl...* (sheep, elephant...),

3. Distinguish words that share the same root but no relationship in the stage of meaning like the noun شجرٌ *šğrun* (trees) and the verb شاجر *šāğr* (dispute),

4. Finally, verify the POS of the rest of the word, since it has an important role in our work.

Most of the Arabic nouns are derived from verbs. Verbs are categorized into their classes. First of all, we see the class of the verb if it is triliteral or not. Classes need to be indicated in each bag of words because different class means different rule to get the correct noun. To better understand the issue let us take a look at the example in table 4 of some verb forms with different classes, their verbal nouns, and examples.

Table 4: Verb forms with verbal nouns and examples.

Verb form	Verbal noun	Example
أَفْعَلَ *ʾafʿala* (a12a3a)	إِفْعَالٌ *ʾifʿaālun* (a12A3u)	أسلم ، إسلام *ʾaslm, ʾislām* (embrace Islam, Islam)
اِنْفَعَلَ *āinfaʿala* (an1a2a3a)	اِنْفِعال *āinfiʿāl* (an1i2A3u)	إنقلب، إنقلاب *ʾinqlb, ʾinqlāb* (Turn over, coup)
فَعَّلَ *faʿala* (1a223a)	تَفْعِيل *tafʿiyl* (ta12I3u)	نفّس، تنفيس *nfs, tnfys* (discharge , discharged)

We can notice that there is a change in verbal noun if we change the class and the form. This issue is detailed with the transformation rules in the next subsection.

5.2 Transformation Rules

As explained before, the rules are the main part of our method because they provide the existence of the relationship and its type. First, the existence of a relation between the pair of words in the same bag is determined by the set of rules in table 5.

Table 5: Transformation rules related to the POS.

№	POS switch	Type of relation	Example
1	Verb → Verb	HasDerivedVerb	أكل، تآكل *ʾakl, tʾākl* (eat, abrade)
2	Verb → Noun	ActiveParticiple	كتب، كاتب *ktb, kātb* (write, writer)
		PassiveParticiple	كتب، مكتوب *ktb, mktwb* (write, written)
		Location	لعب، ملعب *lʿb, mlʿb* (play, stadium)
		Time	غرّب، مغرب *ġrb, mġrb* (go west, Occident)
		Instrument	فتح، مفتاح *fṭḥ, mftāḥ* (open, key)
3	Noun → Noun	HasDerivedNoun	كلب، كليب *klb, klyb* (dog, doggy)
4	Noun → Adjective	Relatedness	سياسة، سياسي *syāst, syāsy* (politic, political)

The problem now is how we can determine the relationship between words in the same bag if it exists of course. Different POS in the same bag is the key for this. Table 5 shows the possible combination in a bag of words that one can find. With the first rules it is easy, if the pair has the same POS (which in this case is a verb) the relation should be *hasDrivedVerb* like the example shows and the same thing goes for the third and the fourth rule. The rule number 2 is a complex one. From all the nouns that you have, e.g you need to distinguish between the active and the passive participle.

The next set of rules will help us to determine all the type of relations between the nouns derived from one verb according to their forms. This will be based on the class of the verb presented in each bag. After a deep look into the behavior of the Arabic verbs and their derivations, the study appears to suggest that we should classify the verbs into two classes, triliteral, and non-triliteral verbs. The table 6 will summarize the transformation rules needed.

Table 6: Transformation rules for the relations between verbs and nouns.

Relation	Verb class	Noun Pattern	Example
ActiveParticiple	Triliteral	فاعل *fāʿl* (1A2i3u)	حمد، حامد *ḥmd, ḥāmd* (praise, praiser)
		weak letter[5] in the 2nd position→ئ *ʾ hamza*	فاح، فائح *fāḥ, fāʾḥ* (spread, Exhalant)
		weak letter in the 3rd position→ي *y ya*	دعا، داعي *dʿā, dʿy* (call, caller)
	Non-triliteral	مُفعِلْ *mufʿil* (mu1a2i3u)	علّم، معلّم *ʿlm, mʿlm* (teach, teacher)
PassiveParticiple	Triliteral	مفعول *mfʿwl* (ma12u3u)	شرب، مشروب *šrb, mšrwb* (drink drinkable)
		م *m* (m)+ the deverbal noun	قال، مقول *qāl, mqwl* (say,)said
	Non-triliteral	مفاعل *mfāʿl* (m1A2i3u)	بارك، مبارك *bārk, mbārk* (bless, blessed)
Location	Triliteral	مفعَل *mfʿal* (ma12a3u)	طبخ، مطبخ *ṭbḫ, mṭbğ* (cook, kitchen)
Time	Triliteral	مفعِل *mfʿil* (ma12i3u)	غرّب، مغرب *ġrb, mġrb* (go west, sundown)

Instrument	-	مفعْل *mfʕl* (mi12a3u)	عول، معول *ʕwl, mʕwl* (count on, pick)
		مفعلة *mfʕlh* (mi12a3h)	قلم ، مقلمة *qlm, mqlmh* (prune, pen case)
		مفعال *mfʕāl* (mi12A3u)	فتح ، مفتاح *fṯḥ, mftāḥ* (open, key)
		فعالة *fʕālh* (1i2A3h)	غسل ، غسالة *ġsl, ġsālh* Washer

To better understand the pattern transformation, you have to think of it as an algorithm. Take the example of the active participle with a triliteral verb who has a weak letter in the second position[6], if such verb does exist in the bag of words alongside with a noun who has a *hamza* in its 3rd position then the relation between them should be made and it is a *activeParticiple* one, and so on for the rest of the nouns. The example of the *instrument* relation, if in the bag of words, a noun with the same pattern as مفعلة *mfʕlh* (mi12a3h) does exist then the relation between its verb should be made.

If you look carefully, the pattern مفعل *mfʕl* (ma12a3u) is presented with four relations, *activeParticiple*, *location*, *time*, and *instrument*. We can distinguish the *activeParticiple* by the diacritics. In our work, the diacritics are token into consideration to affect the proper relations. Beside, AWN's words presented with diacritics. *Location*, *time*, and *instrument* are undistinguished and it is totally logic. The kind of patterns used with those relations are distinguished only in the context. Otherwise, we cannot separate them. Like the words مغرب *mġrb* presented in the example of سافرنا إلى المغرب *sāfrnā ʔilā ālmġrb* (we traveled to Morocco) and عدنا قبل المغرب *ʕdnā qbl ālmġrb* (we come back before sundown) with a different purpose. The first one indicates the location and the second indicates the time. After All these automatic steps we finally can to stage of validation.

5.3 Validation

The steps of the approach are validated according to a lexicographer. The rules too, they are proposed and well studied, as well as the classes of the verbs. Some irregular rules are not taken into considerations because (i) we did not found much of them in AWN or (ii) they will create a confusion with other rules. For example, with nouns, there are other rules like the dual, plural, possessive form. We did not find much of them so we decided to put a general rule for all of them (rule №3 in table 5). We suggested to only work with pertinent rules. We did not go for the automatic validation because the manual verification always leads to better results than the automatic one. It is time-consuming but when you need a better precision you have to sacrifice time.

6 Test and Evaluation

We implemented the method described in the previous section using Java. The first thing we did is cluster words sharing the same root in a bag of words. We notice that some nouns are tagged as an adverb so we verified the POS of each word. Also, some adjectives are wrongly tagged. We corrected as many as we could. We also eliminate named entities and multiword expression because they are *underivatized*. For our own good, The named entities are already tagged so we only eliminated the multiword expression. We only retained nominal, verbal, and adjectival entries. The results are presented in table 7 after the elimination and correction.

Table 7: New frequency of the words in the LMF.

POS	Frequency	New frequency
Noun	16,432	10,325
Verb	42,298	40,143
Adjective	771	498
Total number of bags	6,608	5,462

We fixed the number of bags to 5,462. Each bag has its own set of verbs, nouns, and adjectives and it is cleaned for anything that will misguide the affection of the relation in the next step.

[5]There are 3 weak letters in the Arabic ا، و، ي *ā, w, y* according to their positions in the root we can tell if the verb is *asimilated*, *hollow* or *defective*

[6]This type of verb is called *hollow verb*.

As described in the previous section, the verb class is an important fact in the affectation of the relation. 4,275 bags contain verbs. We classified those bags according to the verb form into two classes. Table 8 shows the detailed frequency.

Table 8: Frequency of verb classes.

Verb class	number of bag of words
Triliteral	3,089
Non-triliteral	1,186

The classification will facilitate the affectation of the relation, which is our next step. All kind of relations described in table 5 was found in the bag of words. Table 9 shows the frequency of each one. Adding the 8,865 new relations to the existing ones, we got 50,001.

Table 9: Frequency of new relations.

Relation	Frequency
HasDerivedVerb	2,005
ActiveParticiple	1,347
PassiveParticiple	1,004
Location	985
Time	752
Instrument	184
HasDerivedNoun	1,784
Relatedness	804
Total	8,865

7 Conclusion

The present paper puts forward a pilot study on the derivational relations between words in Arabic WordNet. Our goal was to engage the specificity of the Arabic word's morphology to enrich the AWN with more precisely relations. Firstly, we clustered the words presented in AWN into a bag of words based on their roots. Secondly, we proposed some morphological rules based on a core of solid linguistic knowledge to identify the existence and the type of relations in each bag of words. Each rule presents the possible patterns that a word can have. Finally, we validated our work with a native speaker and a lexicographer. Our future work will be the test of this new version of the Arabic WordNet in a system like Retrieval Information or Word Sense Disambiguation.

Acknowledgments

This work was supported by the programmer Mr. Abobakr Ahmed Bagais and the lexicographer Mr. Abouloubaba Regui. We would like to thank them for their invaluable advice and encouragement on this research work.

References

Lahsen Abouenour, Karim Bouzoubaa, and Paolo Rosso. 2008. Improving q/a using arabic wordnet. In *Proc. The 2008 International Arab Conference on Information Technology (ACIT'2008), Tunisia, December*.

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.

Rami Ayadi, Mohsen Maraoui, and Mounir Zrigui. 2014. Latent topic model for indexing arabic documents. *International Journal of Information Retrieval Research (IJIRR)*, 4(1):29–45.

Mohamed Ali Batita and Mounir Zrigui. 2017. The enrichment of arabic wordnet antonym relations. In *Computational Linguistics and Intelligent Text Processing - 18th International Conference, CICLing 2017, Budapest, Hungary, April 17-23, 2017*.

Delphine Bernhard. 2010. Apprentissage non supervisé de familles morphologiques: comparaison de méthodes et aspects multilingues. *Traitement Automatique des Langues*, 2(51):11–39.

Orhan Bilgin, Ozlem Cetinoglu, and Kemal Oflazer. 2004. Morphosemantic relations in and across wordnets. In *Proceedings of the Global Wordnet Conference*, pages 60–66.

Mohamed Mahdi Boudabous, Nouha Chaâben Kammoun, Nacef Khedher, Lamia Hadrich Belguith, and Fatiha Sadat. 2013. Arabic wordnet semantic relations enrichment through morpho-lexical patterns. In *Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on*, pages 1–6. IEEE.

Burcu Can and Suresh Manandhar. 2009. Unsupervised learning of morphology by using syntactic categories. In *CLEF (Working Notes)*.

Sabri Elkateb, William Black, Horacio Rodríguez, Musa Alkhalifa, Piek Vossen, Adam Pease, and Christiane Fellbaum. 2006. Building a wordnet for arabic. In *Proceedings of The fifth international conference on Language Resources and Evaluation (LREC 2006)*, pages 22–28.

Christiane Fellbaum, Anne Osherson, and Peter E Clark. 2007. Putting semantics into wordnets "morphosemantic" links. In *Language and Technology Conference*, pages 350–358. Springer.

Christiane Fellbaum. 2010. Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

Nizar Y Habash. 2010. Introduction to arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, 3(1):1–187.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation*, 42(1):21–40.

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.

Svetla Koeva. 2008. Derivational and morphosemantic relations in bulgarian wordnet. *Intelligent Information Systems*, 16:359–369.

Souheyl Mallat, Emna Hkiri, Mohsen Maraoui, and Mounir Zrigui. 2015a. Lexical network enrichment using association rules model. In *CICLing (1)*, pages 59–72.

Souheyl Mallat, Emna Hkiri, Mohsen Maraoui, and Mounir Zrigui. 2015b. Semantic network formalism for knowledge representation: Towards consideration of contextual information. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 11(4):64–85.

Verginica Barbu Mititelu. 2012. Adding morpho-semantic relations to the romanian wordnet. In *LREC*, pages 2596–2601.

Mohamed Achraf Ben Mohamed, Souheyl Mallat, Mohamed Amine Nahdi, and Mounir Zrigui. 2015. Exploring the potential of schemes in building nlp tools for arabic language. *International Arab Journal of Information Technology (IAJIT)*, 12(6).

Karel Pala and Dana Hlaváčková. 2007. Derivational relations in czech wordnet. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies*, pages 75–81. Association for Computational Linguistics.

Maciej Piasecki, Bernd Broda, and Stanislaw Szpakowicz. 2009. *A wordnet from the ground up*. Oficyna Wydawnicza Politechniki Wrocławskiej Wrocław.

Benoît Sagot. 2010. The lefff, a freely available and large-coverage morphological and syntactic lexicon for french. In *7th international conference on Language Resources and Evaluation (LREC 2010)*.

Abdulgabbar Saif, Mohd Juzaiddin Ab Aziz, and Nazlia Omar. 2015. Mapping arabic wordnet synsets to wikipedia articles using monolingual and bilingual features. *Natural Language Engineering*, pages 1–39.

Khaled Shaalan. 2010. Rule-based approach in arabic natural language processing. *The International Journal on Information and Communication Technologies (IJICT)*, 3(3):11–19.

Krešimir Šojat, Matea Srebačić, and Marko Tadić. 2012. Derivational and semantic relations of croatian verbs. *Journal of Language Modelling*, (1):111–142.

Ludovic Tanguy and Nabil Hathout. 2002. Webaffix: un outil d'acquisition morphologique dérivationnelle à partir du web. In *9e Conférence Annuelle sur le Traitement Automatique des Langues Naturelles (TALN-2002)*, page 254.

Delphine Tribout, Anne-Laure Ligozat, and Delphine Bernhard. 2012. Constitution automatique d'une ressource morphologique: Verbagent. In *SHS Web of Conferences*, volume 1, pages 2517–2528. EDP Sciences.

Piek Vossen. 1998. *A multilingual database with lexical semantic networks*. Springer.

W John Wilbur and Larry Smith. 2013. A study of the morpho-semantic relationship in medline. *The open information systems journal*, 6:1.

Wajdi Zaghouani, Abdelati Hawwari, Mona Diab, Tim OGorman, and Ahmed Badran. 2016. Ampn: a semantic resource for arabic morphological patterns. *International Journal of Speech Technology*, 19(2):281–288.