Using the Textual Content of the LMF-Normalized Dictionaries for Identifying and Linking the Syntactic Behaviours to the Meanings

Elleuch Imen
MIRACL Laboratory
B.P. 1088
3018 Sfax, Tunisia
imen.elleuch
@fsegs.rnu.tn

Gargouri Bilel
MIRACL Laboratory
B.P. 1088
3018 Sfax, Tunisia
bilel.gargouri
@fsegs.rnu.tn

Ben Hamadou Abdelmajid
MIRACL Laboratory
B.P. 242
3021 Sakiet-Ezzit Sfax, Tunisia
abdelmajid.benhamadou
@isimsf.rnu.tn

Abstract

In this paper we propose an approach for identifying syntactic behaviours related to lexical items and linking them to the meanings. This approach is based on the analysis of the textual content presented in LMF normalized dictionaries by means of Definition and Context classes. The main particularity of these contents is their large availability and their semantically control due to their attachment to the meanings, which promotes the effective links between the syntactic behaviours and the meaning. In order to test the performance of the proposed approach, we tested it on an available Arabic LMF normalized dictionary. The experiment treats 9,800 verbs and allows us to evaluate the identified syntactic behaviours as well as their links to the meanings.

1 Introduction

A syntactic lexicon is essentially a linguistic resource describing the sub-categorization structure of lexical entries that specify the number and the type of arguments composing the syntactic behaviour. The creation of such a lexicon has been a very large and daunting task. Often, it is approved that the frontier of performance on NLP tasks is shaped entirely by the quality of the syntactic lexicon used. (Carroll and Fang, 2004) showed that the performance of syntactic parsers is improved by using an exhaustive and detailed large lexicon that contains the syntactic knowledge. In the same vein, (Jikoun and Rike, 2004; Surdeanu et al., 2011) argued that a syntactic lexicon represents the core component resource for information extraction, machine translation systems and word sense disambiguation. Due to their importance, several syntactic lexicons appeared for various languages. Regarding English, we can mention FrameNet (Baker et al., 2010), which is a lexical resource for English based on semantic frames and confirmed by attestations in corpus. It aims to document the syntactic and semantic combinatorial (or valence) for each lexical entry through manual annotation of representative lexicographical examples selected from corpus. VerbNet (Kipper et al., 2008) is another lexicon for the English language. It groups verbs sharing the same syntactic and semantic behaviours into classes based on the semantic classification of Levin (1993).

Concerning the French language, we can mention TLFi (Trésor de la Langue Française Informatisé) (Evelyne and Anne-Cécile, 2005), which is a large-scale public resource where sub-categorization is extracted from the dictionary "Trésor de la Langue Française Informatisé". This dictionary, although very structured, was conceived for human use. The lexicon-grammar (Gross, 1975) is another syntactic lexicon for French. It contains information on the syntax of verbs, nouns, adjectives and adverbs into tables.

As regards the Arabic language, we can cite the Arabic VerbNet (Mousser, 2010), which is a syntactic lexicon classifying Arabic verbs into classes based on Levin's verbs classification (Levin, 1993). Another resource for Arabic is ElixirFM (Bielický and Smrž, 2009), which is a functional morphological lexicon enriched with the Arabic verbal frame valence.

All cited lexicons suffer from a problem concerning their models and contents. Thus, such lexicons need to have a large coverage, to guarantee a high level of quality and to be directly usable in NLP tools.

To resolve these problems, the Lexical Markup Framework (LMF) (Francopoulo and George, 2008) ISO 24613 standard has been published providing a convenient solution for the modeling problem. But the enrichment problem still remains. In particular, these lexicons describe the

Proceedings of Recent Advances in Natural Language Processing, pages 168–175,
Hissar, Bulgaria, Sep 7–9 2015.

syntactic behaviours knowledge linked to lexical entries but not to their meanings.

The main goal of this paper is to propose an approach to recognize the syntactic behaviours of lexical entries in LMF dictionaries and to link them to their corresponding meanings. The basic concept of this approach is the analysis of textual contents such as definitions and contexts associated to each meaning of the lexical entries in LMF dictionaries. The main particularity of these contents is their large availability and their semantic control due to their association to the meanings, which promotes the effective links between the syntactic behaviours and the meaning.

The paper is organized as follows: Section 2 presents the proposed approach of self-enrichment of LMF normalized dictionaries with syntactic behaviour linked to the meanings of lexical entries; Section 3 describes our experimentation carried out on an available normalized Arabic dictionary with a discussion of the obtained results; Section 4 exposes related works and their comparison with our study; and finally, Section 5 concludes the paper with the announcement of some future works.

2 Proposed approach

2.1 Fundamentals

The LMF (Francopoulo and George, 2008) provides a standardized framework for the construction of computational lexicons as well as dictionaries for human use. This standard is represented as an object model for structured lexical knowledge by means of a series of extensions (i.e., morphological, syntactic, semantic and syntactico-semantic extensions). In this paper we are interested in the LMF syntactic extension that aims to describe the properties of a lexeme when combined with other lexemes in a sentence. Six classes are reserved to categorize the syntactic descriptions of a lexical entry. The first class is the Sub-categorization Frame that represents one syntactic construction that can be shared by all lexical entry instances. The second class is named the Sub-categorization Frame Set. It represents a set of syntactic constructions and possibly the relationship between them. The Lexeme Property is another class that characterizes one Sub-categorization Frame. Each Sub-categorization Frame is composed of different arguments, represented by the Syntactic Argument class, which allow its connection with the SynSemArgMap instance class. On the other hand, Syntactic Behaviour is the class that describes one of the possible behaviours of a lexeme and it can be attached to the Lexical Entry instance and optionally to the Sense instance.

In an LMF normalized dictionary, a class named Sense is reserved to represent the meaning of a lexical entry. This Sense can be attached to the Definition and Context classes. The Definition class is a narrative description of a Sense. It is reserved for the human user to facilitate his understanding of the meaning. As for the Context class, it represents a text string that describes an example of use of the lexical entry. So, this Context content is displayed for both human use and machine processing.

Benefiting from the particularities of the Context LMF class to be displayed for the computer programs on the one hand, and to describe the uses of the meanings related to the lexical entries on other hand, we propose to analyse this textual content in order to identify the syntactic behaviours of lexical entries then to associate them to the corresponding meanings in LMF normalized dictionaries.

Therefore, the analysis of the Context LMF class related to lexical entries in an LMF normalized dictionaries represents the fundamentals of the proposed approach to identify syntactic behaviours and to associate them to their corresponding meanings.

2.2 Steps of the approach

The proposed approach using the Context of the LMF normalized dictionaries for identifying and linking the syntactic behaviours to the meanings of lexical entries is composed of five steps as shown in Figure 1.

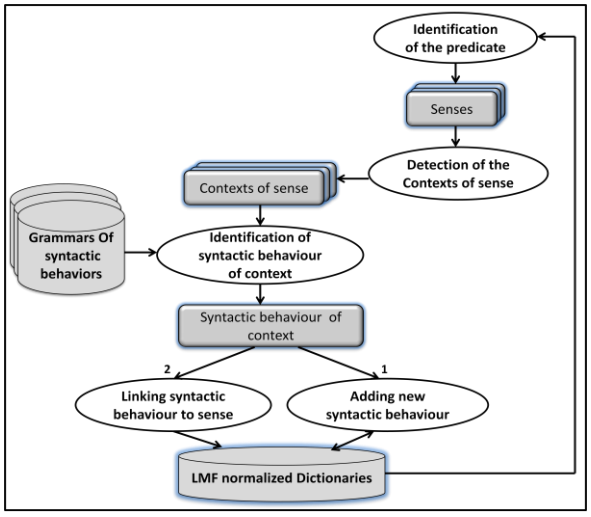

Figure 1: Proposed approach

In the following, we use the verb "to lease", which is extracted from the Oxford Advanced Learner's Dictionary[1] and represented as an LMF lexical entry to detail each step of the proposed approach. As shown in Figure 2 below, this verb has one sense described by four Contexts and one Definition and two syntactic behaviours.

```
▼<Lexicon>
 ▼<LexicalEntry id="53">
    <feat att="partOfSpeech" val="verb"/>
  ▼<Lemma>
     <feat att="writtenForm" val="to lease"/>
   </Lemma>
  ▼<Sense id="53P1">
   ▼<Context>
      <feat att="text" val="We lease all our computer equipment"/>
    </Context>
   ▼<Context>
      <feat att="text" val="They lease the land from a local farmer"/>
    </Context>
   ▼<Context>
      <feat att="text" val="A local farmer leased them the land"/>
    </Context>
   ▼<Context>
      <feat att="text" val="Parts of the building are leased out to tenants"/>
    </Context>
   ▼<Definition>
      <feat att="text" val="to use or let somebody use something, especially
       property or equipment, in exchange for rent "/>
      <feat att="source" val="Oxford Learner's Dictionary"/>
    </Definition>
  </Sense>
  <SyntacticBehaviour id="53C1" senses="53P1" subcategorizationFrames="SVC1C2"/>
  <SyntacticBehaviour id="53C2" senses="53P1" subcategorizationFrames="SVC1toC2"/>
 </LexicalEntry>
```

Figure 2: the verb "to lease" in the LMF dictionary

Identification of the predicate. The role of this step is twofold. Firstly, it searches the predicate to be processed, which can be a verb, an adjective, an adverb or a noun. After that, it aims to find out the meanings represented by the Sense LMF class attached to the processed predicate.

The application of the first step on the example presented in Figure2 identifies the predicate having 53 as identifier and "to lease" as lemma. One sense marks this predicate identified by the identifier "53P1", which corresponds to the first principal meaning of the "53" lexical entry in the LMF dictionary.

Detection of the Contexts of sense. A Context LMF class is used to describe the use of the lexical entry by means of a simple sentence. These Contexts are marked by their broad availability in the dictionary and by their semantic endorsement due to their association with the meanings. In order to find out syntactic behaviours and to link them to Senses, we propose to analyse these Contexts. Thus, the purpose of this step is to search for the processed sense related to lexical entry all linked Contexts.

For the Sense "53P1" related to the verb "to lease", the second step of the proposed approach identifies four Contexts: (1) "We lease all our computer equipment", (2) "They lease the land from a local farmer", and (3) "A local farmer leased them the land" and (4) "Parts of the building are leased out to tenants".

Identification of the syntactic behaviour in Context. This step aims to identify the syntactic behaviour for each Context recognized in the previous step. To accomplish this objective, this step uses Grammars of syntactic behaviours. These Grammars must be constructed by means of linguistic tools and must be able to put a sentence in input in order to recognize its corresponding syntactic behaviour. At the end of this step, for each processed Context the syntactic behaviour is identified.

When we applied the third step to the Contexts obtained previously, we obtained the results described below. For the first context, "We lease all our computer equipment", Grammars of syntactic behaviour parses this sentence and recognizes the following: "We": the Subject, "lease": the processed predicate and "all our computer equipment": the Object. So, the corresponding syntactic behaviour is SVC (Subject Verb Complement). For the second Context, the SVC1fromC2 (Subject Verb First Complement "from" preposition Second Complement) syntactic behaviour is identified. Concerning the third Context, its related syntactic behaviour is SVC1C2 (Subject Verb First Complement, second Complement). As regards the fourth Context, the Grammars of syntactic behaviours identify the SVC1toC2 (Subject Verb First Complement "to" preposition Second Complement) syntactic behaviour.

Adding new syntactic behaviour. In the LMF normalized dictionaries, an existing list of syntactic behaviours can be linked to lexical entries, whereas the application of Grammars of syntactic behaviours to Contexts can identify new syntactic behaviours that do not appear in this list. At this stage, these new syntactic behaviours must be added to the list of syntactic behaviours related to the processed predicate.

Two syntactic behaviours, namely SVC1C2 and SVC1toC2, are linked to the predicate of the verb "to lease" in the example of Figure2. The application of Grammars of syntactic behaviours to Contexts identifies two new syntactic behaviours: SVC and SVC1fromC2. These later

[1]http://www.oxfordlearnersdictionaries.com/definition/english/lease_2

will take two new identifiers “53C2” and “53C3” having the sub-categorization Frames respectively: SVC and SVC1fromC2.

Linking syntactic behaviour to sense: At this stage, we have a final list of syntactic behaviours related to the processed lexical entry. Then, the objective is now to associate each syntactic behaviour to its corresponding Sense meaning.

For the verb “to lease”, all syntactic behaviours whatsoever, already existing or identified by the application of Grammars of syntactic behaviours, are related to the “53P1” sense. Thus, for each Syntactic Behaviour class an attribute named sense will be added having the value “53P1”.

3 Experiment and results

To consolidate our proposed approach, we tested it on an available Arabic LMF normalized dictionary. So, in this section we will present the available Arabic dictionary with its component knowledge. Then, we will detail the experimentation carried out and comment on the obtained results.

3.1 The LMF normalized Arabic dictionary

An Arabic LMF normalized dictionary named El-Madar[2] has been developed by (Khemakhem et al., 2013). The model of this dictionary takes into account the specificities of the Arabic language and covers the morphological, syntactic, semantic and syntactico-semantic levels. The current version of this dictionary contains about 37,000 lexical entries: 10,800 verbs, 22,400 nouns and 3,800 roots. Each lexical entry can include a morphological content like the part-of-speech, the lemma, some derived and inflected forms, etc. Also, it contains semantic knowledge such as the synonymy that can join senses of entries. Concerning the syntactic content, the El-Madar dictionary contains 155 general syntactic behaviours related to Arabic verbs where 5,000 verbs are connected to those behaviours.

3.2 The experiment

Our experimentation uses the El-Madar Arabic LMF dictionary. We are limited in this paper to processing verbal predicates. Apart from that, each step of the proposed approach will be experimented on the verbal predicate “وَهَبَ/ wahaba / to give” derived from El-Madar dictionary.

Experimentation of the “identification of the predicate” step. Figure3 presents the experimentation of the identification of the predicate step applied to the verb “وَهَبَ/ wahaba / to give”.

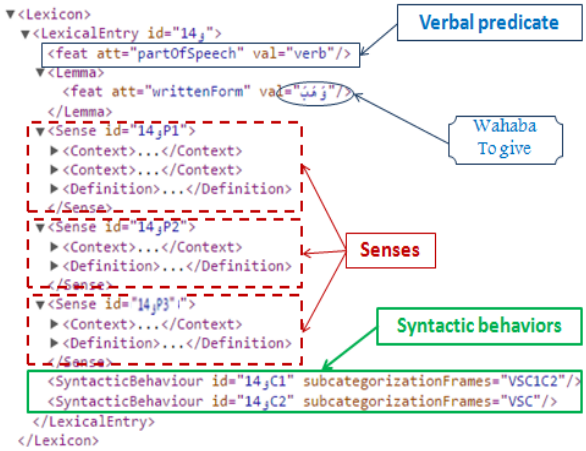

Figure 3: Experimentation of the identification of the predicate step

The lexical entry in Figure3 corresponds to the verbal predicate having the lemma “وَهَبَ[3]/wahaba/to give” and the identifier id=”14و”. This verb has three senses identified respectively “14وP1”, “14وP2” and “14وP3” and two syntactic behaviours “14وC1” and “14وC2”. This first step aims to recognize this verbal predicate.

Experimentation of the “detection of the contexts of sense” step. The same verbal predicate “وَهَبَ/wahaba/to give” is used at this stage to experiment the detection of the contexts of sense step. Figure 4 details this experimentation.

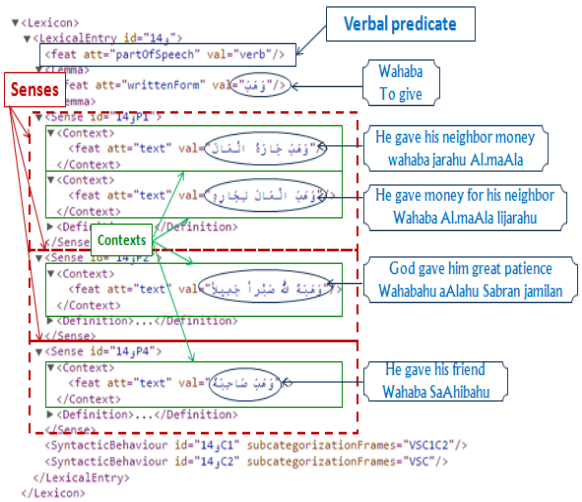

Figure 4: Detection of contexts of sense of the verb “وَهَبَ/wahaba/to give”

[2] http://elmadar.miracl-apps.com/

[3] the Arabic transliteration which has been used is Habash, Soudi and Buckwalter (Habash et al, 2007)

The experimentation of the second step on the verbal predicate "وَهَبَ/wahaba/to give" can recognize two contexts related to the sense1 id="14وP1": "وَهَبَ جَارَهُ الْمَالَ/wahaba jarahu Al.maAla/ He gave his neighbor money" and " وَهَبَ الْمَالَ لِجَارِهِ/ wahaba Al.maAla lijarihi/He gave money to his neighbor". For the second sense id="14وP2" one context is identified " وَهَبَهُ اللَّهُ صَبْراً جَمِيلاً/wahabahu Aalahu Sabran jamilan/ God gave him great patience". Regarding the third sense id="14 و P3" the only context found is " وَهَبَ صَاحِبَهُ /wahaba SaAhibahu/He gave his friend".

Experimentation of the "identification of syntactic behaviour of context" step. After searching contexts for each sense of the lexical entry "وَهَبَ/wahaba/ = to give", the identification of corresponding syntactic behaviours takes place.

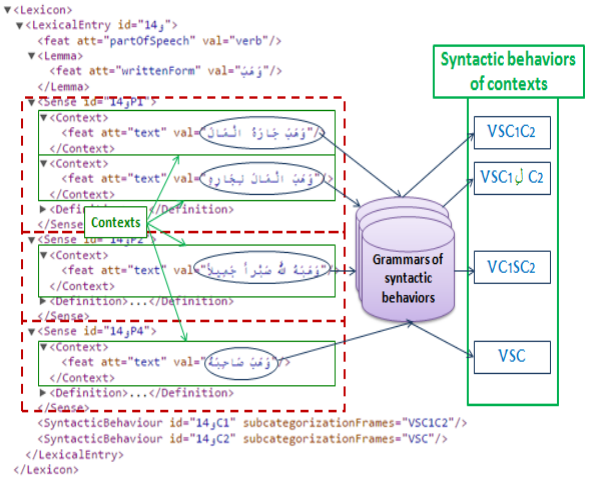

Figure 5: Experimentation of the "identification of syntactic behaviour of context" step

Figure5 demonstrates the recognition of syntactic behaviours of contexts of the verbal predicate "وَهَبَ/wahaba/to give". This identification is realized by the Grammars of syntactic behaviours. Those grammars (Elleuch et al., 2013) have been constructed using the NooJ[4] linguistic platform according to all existing Arabic syntactic patterns. They are able to identify for a simple sentence in input its corresponding syntactic behaviour. For example, when we applied Grammars of syntactic behaviours to the context " وَهَبَهُ اللَّهُ صَبْراً جَمِيلاً/wahabahu Aalahu Sabran jamilan/God gave him great patience" of the sense id="14وP2", the result of this application is VC1SC2. Indeed, the grammar parses the context in tokens: "وَهَبَهُ", "اللَّهُ", "صَبْراً" and "جَمِيلاً". The grammar can recognize "وَهَبَهُ/wahabahu/ gave him" as an agglutinate token composed of "وَهَبَ/wahaba/give", which is the verb (V), and "هُ/hu/him", which is a pronoun agglutinate to the verb representing the first complement (C1). "اللَّهُ/Aalahu/God" is a noun that fulfils the function subject (S). "صَبْراً/Sabran/patience" is a noun and "جَمِيلاً" is an adjective that describes "صَبْراً/Sabran/patience"; thus, "صَبْراً جَمِيلاً/Sabran jamilan/great patience" satisfies the function of second complement (C2).

The application of Grammars of syntactic behaviours to contexts finds the syntactic behaviours VSC1C2 and VSC1لِC2 for Sense1. The syntactic behaviour VC1SC2 is identified for the context of the second sense. Also, the syntactic behaviour VSC is recognized for Sense3 of the treated lexical entry "وَهَبَ".

Experimentation of the "addition of a new syntactic behaviour" step: Figure 6 below illustrates the experimentation of the enrichment of the "addition of a new syntactic behaviour" step.

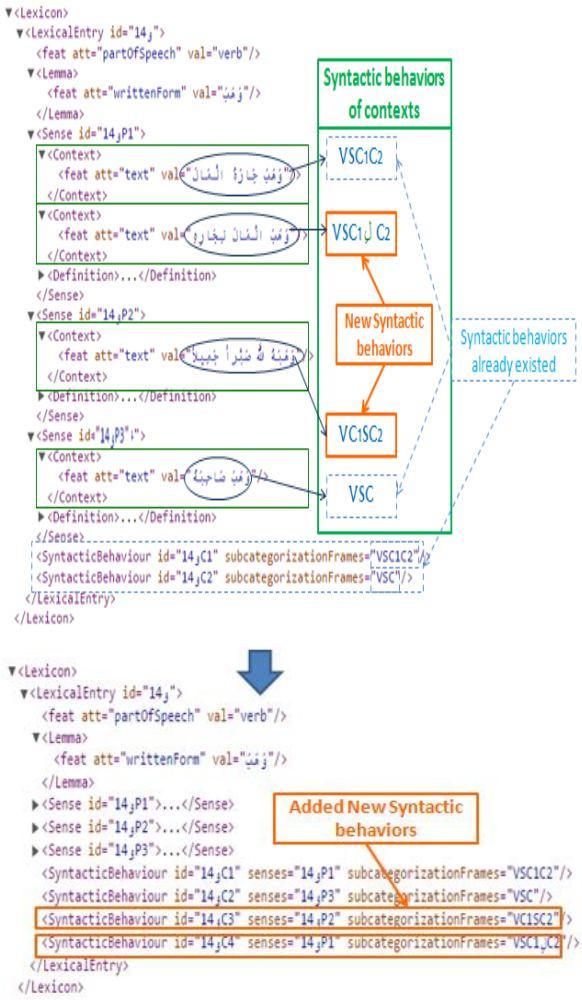

Figure 6: Adding new syntactic behaviour experiment

[4] www.nooj4nlp.net

As illustrated in Figure 6, this step makes a comparison between the already existing syntactic behaviours with the syntactic behaviours identified in the previous step. Indeed, when we compare the syntactic behaviours related to the predicate "وَهَبَ" with the syntactic behaviours identified for the contexts, we note that VSC1لِC2 and VC1SC2 are newly detected syntactic behaviours. Then, the "addition of new syntactic behaviour" step appends those new syntactic behaviours to the predicate id="14و" "وَهَبَ". In this stage, the predicate "وَهَبَ" has four syntactic behaviours: VSC1C2, VSC, VSC1لِC2 and VSC1C2.

Experimentation of the "linking syntactic behaviour to sense" step. The experimentation of the "linking syntactic behaviour to sense" step is presented in Figure7.

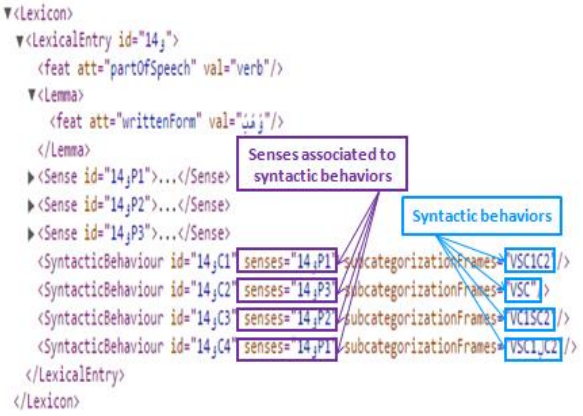

Figure 7: "Association of syntactic behaviour to sense" experiment

Figure7 represents the addition of the identifier of sense to each syntactic behaviour.
As VSC1C2 and VSC1لِC2 are identified in the first sense, the identifier id="14وP1" of this sense is added to the syntactic behaviours VSC1C2 and VSC1لِC2. Since the syntactic behaviour VC1SC2 is recognized in the context of the second sense, the id="14وP2" of the second sense is added to the syntactic behaviour VC1SC2. And finally, the id="14 و P3" of the third sense will be associated to the syntactic behaviour VSC where this behaviour is identified in the context of this sense.

3.3 Results

El-Madar dictionary (Khemakhem et al., 2013) contains up to now 10,800 verbs. Among them 1,000 verbs don't have the Sense classes. So, only 9,800 verbs have been treated by the experimentation we performed. 31,500 assignments between syntactic behaviours and meanings are the result of the experimentation of the proposed approach applied to El-Madar dictionary. A sample containing 2,000 resulting affectations representing the 155 kinds of Arabic syntactic behaviours have been assessed by a human expert. For these 2,000 affectations, the expert approves that 232 incorrect affectations and 140 missed ones are detected. Thus, for these 2,000 affectations the Precision is estimated to 0.88 and the Recall is equal to 0.92.
For error analysis, we can acknowledge that the sentence of the processed Context is represented as a complex structure and the Grammars of syntactic behaviours cannot analyse it and give wrong results. Also, we can accept that the Context written by the lexicographer is not appropriate to the exact syntactic behaviour of verbs.

4 Related works

In this section, we will present an overview of some Arabic syntactic lexicons. We can mention the ElixirFM lexicon (Bielický and Smrž, 2009), the Arabic syntactic lexicon (Loukil et al., 2010), and the Arabic VerbNet (Mousser, 2010) syntactic lexicons for the Arabic language since we have experimented the proposed approach on this language. At the end of this section, we will make a comparison between the three mentioned lexicons with our lexicon.

4.1 The ElixirFM Lexicon

ElixirFM (Bielický and Smrž, 2009) is a morphological lexicon enriched by the valency frame of Arabic verbs. This lexicon is based on the theoretical Functional Generative Description (FGD) approach. The valence of a verb is represented as a tree of dependencies. The lexicon contains about 3,500 frames of verb valence: 2,000 frames representing the intransitive verbs automatically created from the Buckwalter Arabic Morphological Analyzer and 1,500 frames manually formed. These frames take into account the thematic role of each argument which is composed of the syntactic behaviours of Arabic verbs and which also includes both obligatory and optional actants and only obligatory free modifications. In fact, this lexicon does not take into consideration the valency of modal, impersonal and defective verbs.

4.2 The Arabic Syntactic Lexicon

The Arabic syntactic lexicon (Loukil et al., 2010) is a lexical resource compliant to the LMF standard representing the syntactic features of Arabic verbs. The enrichment process used to populate this resource with syntactic behaviours is made semi-automatically by means of the editor Lexus. Three steps compose the enrichment process. The first step is the manual identification of syntactic behaviours for Arabic verbs. The second one represents the use of the Lexus editor in order to enrich the lexicon with sub-categorizations of verbs. The last step details how to edit and affect sub-categorization frames to each processed verb. This lexicon includes 2,500 verb lemmas.

We can mention that the Arabic syntactic lexicon doesn't cover all syntactic behaviours of Arabic verbs because it considers only 17 sub-categorization frames. Also, the affectation of the sub-categorization frames is attached to the lexical entry but not to its meanings.

4.3 The Arabic VerbNet

The Arabic VerbNet (Mousser, 2010) is the Arabic version of the English VerbNet. It is a lexicon that classifies Arabic verbs based on Levin's classification (Levin, 1993). Thus, the same procedure, process and treatment used to build the English VerbNet were re-used to construct the Arabic VerbNet, with some adaptation for the Arabic language. This lexicon classifies verbs into classes. Each class groups verbs sharing syntactic and semantic properties represented into frames. Morphological, syntactic and semantic knowledge are presented into each frame. Indeed, the root, the derived forms, the present participle of the Arabic verb, the thematic roles of semantic arguments and the sub-categorization of each verb are included into each frame. 291 is the number of verb classes of the Arabic VerbNet including 7,937 verbs represented with 1,202 frames.

4.4 Synthesis

Even though all the approaches presented in the above studies on the Arabic language suggest some interesting ideas, each one of them includes some shortcomings. Indeed, ElixirFM does not present the explicit syntactic structure of verbs and neglects the syntactic functions of complements. The syntactic lexicon of (Loukil et al., 2010) is a very small lexicon representing only the syntactic aspects of very few Arabic verbs while the Arabic VerbNet does not represent the native features of Arabic verbs because it's a simple translation of the classes used in the English VerbNet with some adaptations.

A comparison between those three works and our lexicon according to different criteria is presented in Table 1, which is given below.

	ElixirFM Lexicon	**Arabic Syntactic Lexicon**	**Arabic Verbnet**	**Our Lexicon**
Normalized format	-	**LMF standard**	-	**LMF standard**
Linguistics levels covered				
Morphology	+	+	+	+
Semantic	+/-	-	+	+
Syntactic	+	+	+	+
Number of verbs	3500	2500	7937	9800
Classification	FGD	Verbal type hierarchy	Levin's class	Transitivity Intransitivity
Syntactic enrichment				
Process	Semi-automatic	Semi-automatic	Semi-automatic	Automatic
Formal representation	Frames	32 Rules	English Verbenet Frames	155 Grammars
Semantic features	+	-	+	-
Dictionaries	Printed	-	Printed	Electronic
	3		2	Arabic normalized LMF dictionary
Corpus	PADT	205 transliterated sentences	-	-
	CLARA			
	Arabic Gigaword			
Syntactic behavior				
Number	3500	17	1202	155
Related to	Meanings	Verbs	Meanings	Verbs Meanings

Table 1: Comparison with the existing Arabic syntactic lexicons

5 Conclusion and perspectives

We have presented an approach allowing us to find out the syntactic behaviours of lexical entries and linking them to their corresponding meanings in LMF normalized dictionaries. This approach uses the Context textual content to identify the syntactic behaviour. The main particularity of this content is its large availability and its semantic control due to this connection to the meanings, which promotes the effective links between the syntactic behaviour and the meaning. This approach is characterized by it genericity; thus it can be applied to any language. We have tested the proposed approach by its application to the Arabic language. For that purpose, an available Arabic LMF normalized dictionary named El-Madar was used to evaluate our approach. 9,800 verbs were treated in the experimentation giving 0.88 of Precision and 0.92 of Recall.

Future directions include extracting syntactic behaviours from other resources like corpora,

and improving Grammars of syntactic behaviours in order to make them more sophisticated to support more complex linguistic rules.

References

Baker K., Bloodgood M. , Chris D., Bonnie W., Filardo N., Levin L., Miller S. and Piatko, C. 2010. *Semantically-Informed Machine Translation: A Tree-Grafting Approach*. Biennial Conference of the Association for Machine Translation in the Americas. Denver, Colorado, pp.411-438.

Bielický V. and Smrž O. 2009. *Enhancing the ElixirFM Lexicon with Verbal Valency Frames*, volume 2. International Conference on Arabic Language Resources and Tools (MEDAR 2009), Cairo, Egypt.

Carroll J., Fang A.C. 2004. *The Automatic Acquisition of Verb Subcategorisations and their Impact on the Performance of an HPSG Parser*, volume1. International Conference on Natural Language Processing, Sanya City, China.

Elleuch I., Gargouri B. and Ben-Hamadou, A. 2013. *Syntactic enrichment of Arabic dictionaries normalized LMF using corpora*. Language & Technology Conference (LTC). Poznan, Poland. pp.314-318.

Evelyne J. and Anne-Cécile N. 2005. *Vers un Lexique Syntaxique du Français : Extraction d'Informations de Sous-Catégorisation à Partir du TLFi.* Journée ATALA du 12 mars sur les lexiques syntaxiques, Paris.

Francopoulo G. and George M. 2008. *ISO/TC 37/SC 4 Rev.16. Language Resource Management − Lexical Markup Framework (LMF).*

Gross M. 1975. *Méthodes en Syntaxe: Régimes des Constructions Complétives*. Paris, Hermann.

Habash N., Soudi A. and Buckwalter T. 2007. *In Arabic Computational Morphology: Knowledge-based and Empirical Methods.* ISBN: 978-1-4020-6045-8.

Jikoun J.M. and Rike M. 2004. *Information Extraction for Question Answering: Improving recall Through Syntactic Patterns*, volume20. International Conference on Computational Linguistics

Khemakhem A., Gargouri B. and Ben Hamadou A. 2013. *LMF for Arabic, chapter in the book "LMF : Lexical Markup Framework".* TWiley Edictions, ISBN: 9781848214309, pp. 83-96.

Kipper K., Korhonen A., Ryant N. and Palmer, M. 2008. *A Large-Scale Classification of English Verbs*, volume 42. Journal of Language Resources and Evaluation, n° 1, pp. 21-40

Levin B. 1993. *English Verb Classes and Alternations*. Preliminary investigation, University of Chicago Press, Chicago and London.

Loukil N., Haddar K. and Ben Hamadou A. 2010. *A Syntactic Lexicon for Arabic Verbs.* International Conference on Language Resources and Evaluation, LREC, pp. 17-23, May, Valletta, Malta.

Mousser J. 2010. *A large Coverage Verb Taxonomy For Arabic.* International Conference on Language Resources and Evaluation, LREC, May, Valletta, Malta.

Surdeanu M., McClosky D., Smith M-R., Gusev, A. and Manning C-D. 2011. *Customizing an Information Extraction System to a New Domain.* In Proceedings of the ACL 2011 Workshop on Relational Models of Semantics (RELMS 2011), Portland, Oregon, USA, June 23, 2011. pp. 2–10.