Building a Lexical Semantic Resource for Arabic Morphological Patterns

Abdelati Hawwari
Columbia University
New York City, USA
ah3019@ columbia.edu

Wajdi Zaghouani
Carnegie Mellon University
Doha, Qatar
wajdiz@qatar.cmu.edu

Tim O'Gorman
University of Colorado
Boulder, USA
timothy.ogorman@colorado.edu

Ahmed Badran
Fayoum University
Fayoum, Egypt
afz00@fayoum.edu.eg

Mona Diab
George Washington University
Washington, DC, USA
mdiab@gwu.edu

Abstract—We present a pilot study for building an Arabic morphological Pattern Net as a lexical semantic resource. In this paper, a limited number of Arabic Morphological Patterns have been selected in order to analyze the structure and behavior of some verbs in the Arabic PropBank [19]. Our goal is to study whether there is a direct relationship between morphological patterns and verbal semantic roles. The approach to building our morphological patterns' database is based on linguistic generalization of the semantic roles of the verbal predicates. The results obtained show the feasibility of a more comprehensive future study.

Keywords— Arabic Morphology; Lexical-Semantics; PropBank

I. INTRODUCTION

Arabic morphology exhibits both derivational and inflectional morphology. The derivations are known for their templatic structure. The templates are composed of a combination of roots and patterns, with some degree of idiosyncracy. Roots encode semantic information and they combine with patterns, which encode some syntactic information. The roots are sets of either three or four consonants, referred to as radicals, which are inserted at specific points within a morphemic pattern. For example, the root k–t–b, "the act of writing", could be substituted into the numbered radical slots of a pattern [c¹ac²c²ac³a] to form a lexeme, as shown in the figure 1 below:
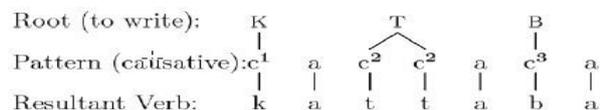
Figure 1. How roots and patterns form lexemes

Derived words, or derivations such as *kat~ab*, are the lexemes that are further inflected for number, gender, case, mood, aspect, definiteness, etc., to reflect the inflectional morphological system. In general a specific lexeme is chosen conventionally to be the citation form for the word and it is referred to as the lemma form, for verbs it is the 3rd person masculine singular form, for example.

Verbal pattern morphemes have a set of related 'deverbal' morphemes encoding a variety of syntactic relationships of the term, such as 'place of verb', 'instrument of event'. Thus a single derived form of a verb, such as the combination of the root f-t-H, "to open", and a pattern [ic¹tac²ac³a], may result in all the deverbals in the Table 1 below. The details of how roots and patterns combine are valid questions, ones already thoroughly addressed in work on morphological analysis such as [22], [12], and [7]. This lexical resource therefore will focus entirely upon the semantic and syntactic consequences of these patterns.

TABLE I.   DEVERBALS DERIVED FROM A SINGLE COMBINATION OF ROOT (FTH) AND PATTERN (IC¹TAC²AC³A).

AifotataHa (*open, inaugurate*)				
MASDAR	ACTIVE	PASSIVE	PLACE	TIME
AifotataH	mufotatiH	mufotataH	mufotataH	mufotataH
Opening, beginning	*one who opens*	*the thing inaugurated*	*place of opening*	*time of opening*

In this pilot study, a limited number of Arabic Morphological Patterns have been selected in order to analyze the structure and the behavior of the verbs in the Arabic PropBank [19], which is a semantically annotated corpus. Our goal is to find the relationship between morphological patterns and the verbal semantic roles. In the next section we will present the background on some of the published related research in Arabic Morphology. Section 3 will provide the motivation and the merit behind this pilot study. Section 4 will provide a detailed description of our methodology with the details of the various steps involved in creating this pilot resource.

II. RELATED WORK

It is worth noting that non-computational classifications of the Arabic morphological pattern system have been investigated in [13], [8], [3] and [23]. Many of the studies in English like in [23] have focused only on a ten core ubiquitously taught trilateral patterns - patterns for roots composed of three radicals. These studies are incomplete in their coverage since there exist quadrilateral roots and even rare 5-radical roots, moreover within the trilateral root pattern template system, there are more than ten trilateral templates. However, more importantly, while many studies have casually noted various meanings for the patterns, none of these studies did so at the fine level of granularity we are proposing or even in a systematic formal manner, but rather simply assumed a singular conceptual structure for each pattern. Similarly, the work of [13] investigates the classes and syntactic alternations of Arabic verbs and adjectives, yet it does not distinguish between the inflectional verbal patterns and other derivational morphemes, hence mixing two separate morphological operations (inflection and derivation), masking the regularities that exist latently in the pattern morphology. In contrast, Almutawakkal developed in [3] a coherent classification of Arabic morphological derivation using the Functional Grammar framework, classifying Arabic verbs according to their argument structure characteristics and derivational patterns. Almutawakkal's work, however, is entirely theoretical, and therefore, neither grounded in empirical data extracted from a corpus, nor computationally accessible. Yet it provides a useful set of descriptions against which this pilot study's

conclusions could be compared. Accordingly, we are planning to use it as a primary reference point for our proposed investigation.

Although there exist a variety of lexical resources for Arabic, few provide a systematic investigation of the semantic/syntactic aspects of the morphological system. Resources for morphological analysis, the underlying represented lexica, such as those used in the following systems: [5], [6], [7], [9] and [16] allow for the analysis of words into roots and patterns, yet when they provide semantic information via explicit semantic fields, the granularity is at the level of the verb lemma. We believe this to be much too coarse grained to exhibit the underlying systematic regularity. The most morphologically detailed of these systems, the system in [5], treats the morphological pattern as one possible feature for determining a semantic field of a new term. Arabic Wordnet similarly contains information about semantic relations, but that is also encoded at the lemma level without addressing roots and patterns systematically.

Two more syntactically grounded resources, more relevant to the task at hand, are the Arabic PropBank [19] and [24] and Arabic VerbNet [18]. Arabic PropBank is a corpus of treebanked sentences in which the predicates have been disambiguated and their argument structure labeled. Arabic VerbNet [18], based upon the English VerbNet project [15], is a mapping of Levin's syntactic alternations into Arabic, with alternations added whenever necessary. This provides a rough verbal classification of syntactic alternations, with the same semantic and syntactic representations of these classes as utilized in English VerbNet.

One complexity of Mousser's Arabic VerbNet, however, arises from the richness and power of the morphological system itself. Syntactic alternations in English – such as the difference between "John dried the clothes" and "the clothes dried" – may be expressed in Arabic using different, yet very similar verb lemmas - in this instance, na$afa and na$$afa respectively (example from [18]:1). As such, Mousser introduces "sibling classes", essentially treating certain morphological variants as part of the same syntactic alternation [18]. Hence it is on the one hand fine grained in representing shared syntactic and semantic information across verb lemmas, yet coarse grained in grouping some of these lemma templates. Crucially however, we believe, that this exemplifies the complexity of dealing with such a morphologically-rich language without having a core resource systematically modeling the intricacies of the underlying morphology.

These lexical resources have, however, given Arabic NLP enough access to the Arabic morphological system - albeit in perhaps a simplified manner - to allow basic forays into its use. Explicitly modeling aspects of the morphological system in NLP has led to improvements in performance even when using simplistic representations of the root pattern system or explicit encodings of morphological features as evidenced in tasks such as Information Retrieval [4] and Arabic Semantic Role Labeling [10]. Such NLP systems would presumably see even better improvements if they had access to richer lexical resources for Arabic morphology, especially if there existed a VerbNet style database explicitly encoding syntactic regularities at the pattern sense level of granularity proposed here. Counter intuitively, several systems that attempted to explicitly represent a 'pattern' feature, given the current resources, failed to show relative improvement in performance. For example, the automated verb classification system described in [20] and the dependency parsing of [17] found no improvements in the results despite the inclusion of an explicit 'pattern' feature. This may be because each disambiguated pattern packs a vast range of syntactic alternations that are not correctly identified for such systems. The level of granularity for the pattern encoding was too coarse grained. The results of these systems findings are not discouraging, but rather evidence that the morphological system of Arabic is too complicated to approximate without semantically rich lexical resources. A system such as [20] would have benefited tremendously from a computationally accessible pattern sense inventory with associated consistent sets of syntactic alternations.

III. MOTIVATION BEHIND THE STUDY

The Arabic Morphological Pattern Net (AMPN) aims at representing both the complex, derivational structure for the morphological root patterns of Arabic, while also dealing with the multiple senses within each of these patterns in turn. These morphological roots (which carry the semantic content) combined with the patterns that regulate the syntactic clausal behavior of the lexicon) added together provide the 'lemma' of the verb. Each verb lemma, in turn, would be marked for the particular sense of its derivational pattern. In this respect, it is presumed that a system addressing these issues would base its conclusions upon actual verbs in context. It is further noted that the regularities within each sense noted above all concern issues of semantic form, selectional restriction, and argument structure. A resource representing these senses and their derivations would thus need to represent these regularities in a consistent manner, such as the one used in systems like VerbNet [15]. Such a resource could be valuable for Arabic NLP and could help address undisambiguated pattern feature that has been useful to date in Arabic NLP tasks. This, in turn, would enhance the robustness of Arabic NLP when it comes to the analysis of the Arabic morphological system that has always been a major challenge for natural language processing [21].

IV. METHODOLOGY

The following assumptions underline our choice of representation:

(1) Morphological patterns, though non-contiguous, can be handled like morphemes, and given lexical entries as such;

(2) These morphemes are polysemous, and their different senses will display different syntactic alternations;

(3) Each sense will be represented using a list of which verb roots combine with that sense.

The suggested system of analysis depends on lexical entries of verb lemmas that consist of a combination of roots and morpheme patterns rather than individual words. These lexical entries are analyzed along a number of dimensions, to capture the various semantic, syntactic, and semantic layers of the pattern. Accordingly, we started from APB [19].

To further investigate our Assumptions, we selected 2100 APB verb frames and looked at their application to ten specific morphological Patterns. The selected patterns represent three levels of the assumed derivational hierarchy as follows: **(1)** C1aC2C3aC4, **(2)** C1aC2C2aC3, C1aaC2aC3, CaC1C2aC3, C1taC2aC3, nC1aC2aC3, and stC1C2aC3, **(3)** taC1aC2C2aC3, taC1aaC2aC3, and taC1aC2C3aC4). Firstly, we annotated the APB verb file (2100), added a Morphological Pattern and Root to each verb (verb lemmas), and made a sub file for each different Morphological Pattern (MP) that can form a verb.

For each MP, we grouped its possible frames of analysis using many factors that include: (1) the syntactic alternations of the verb, (2) number of arguments (3) nature of arguments, (4) the semantic features of the arguments, (5) the semantic Roles, and Semantic Classification, as well as (6) Aspectual Types. The verb can exist within more than one frame of analysis. Accordingly, we adopted some verb syntactic characteristics to classify the Arabic verbs based on their syntactic alternations. This, in turn, allowed us to capture the different senses that each morphological pattern may have. An example is illustrated in Figure 2, which shows all of the lemmas formed by a single root, f-t-H:

f-t-H   (**Roots**)
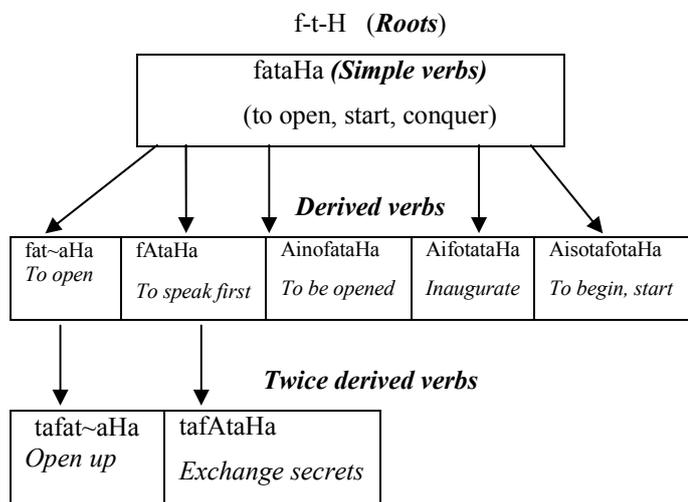
Figure 2. The derivation of all verb lemmas for the root f-t-H

In depending on the APB, we employed two major features: Firstly, the argument structure of the morphological pattern; that is, distinguishing the number of arguments for each verb (ARG0 or ARG1). In addition, we employed the lexical representation of the APB to explore the nature of the verb arguments by exploring the semantic roles as well as the semantic features of arguments in terms of 'rationality' (Rational/ Irrational/ Both); and number (Singular/ Plural/ Both) with the exception of the simple trilateral verb.

A. Annotation Methods

The details of how we implemented our proposal involve three different kinds of annotation that have some interdependencies: (a) clustering predicate PropBank senses into specific senses for specific morphological patterns; (b) developing detailed lexical entries for each sense in the created clusters (it is worth noting that each of these clusters is linked to numerous verbal and deverbal entries via PropBank); (c) linking these senses to each other, so as to model semantic derivation rules, shared syntactic alternations

for the different patterns to discover possible derivational gaps.

B. Clustering into different senses

The first task of clustering predicate PropBank senses into specific senses for specific morphological patterns begins with a quick annotation of all predicate instances within the Arabic PropBank corpus for their morphological patterns. We have done this task manually for our 10 verbs, but it is relatively simple task, which may be preprocessed later on using some automatic taggers using morphological disambiguators such as MADA [14], or AMIRA [11], followed by a stage of manual correction.

C. Lexical Entries for each sense

The second group of tasks is the creation of detailed lexical entries for these senses. Again, done manually for our ten verbal patterns. This has some similarity to verb classification systems such as VerbNet [15] and Arabic VerbNet [18]. Therefore a roughly similar formalism is used for the main elements of these entries; the description of syntactic alternations, the expression of thematic roles, and the use of semantic notation to express the semantics of a pattern.

The annotation of syntactic alternations is based upon sentences in the Arabic PropBank corpus, which are examined to reveal the syntactic features of all the verbs in a given sense. As each sense of a pattern is presumed to be syntactically coherent, each set of alternations should be different than those of the other senses. The vocabulary of roles used in this annotation is the same set of 30 thematic roles used in the current VerbNet project [15], and the same notation of syntactic restrictions (currently 41).

The syntactic alternations follow VerbNet conventions with the exception of the Arabic 'subject', which can have a number of positions, none of which seem to be determined by verbal syntax. Thus subject position is always treated as being in the canonical VS(O) position after the verb, which parallels the treatment of Arabic Treebank, by treating SVO position as topicalized. The resultant form of syntax, for something such as the reflexive form of [ic¹tac²ac³a], shows:

(1) Verb Actor (Agent+Patient) <+plural>

(2) Verb Actor (Agent+Patient) <+plural> {preposition} Topic

We also propose annotating for grammatical aspect as implied by the pattern. Each pattern conveys not simply valence information, but also specific grammatical aspect, with many senses of the simple morphological patterns roughly corresponding to Vendler aktionsart classes ([C¹aC²uC³-u] being generally stative, [C¹aC²iC³-a] being accomplishments, etc) and more complicated patterns involving other aspects, so that [taC¹aaC²aC³a] may involve gradual repeated events (such as *tatabaE*, meaning to follow one another consecutively).

The relationship between the annotation of the grammatical aspect of these senses, and the semantic representation of the same senses is kept separate (despite the theoretical capacity of a semantic representation such as that of VerbNet to handle aspectual complexity) so as to avoid complexities involving how grammatical aspect is inherited.

Finally, using a conceptual semantic structure similar to VerbNet, models for the semantics of a given pattern are also presented. However, unlike verbal classifications which presume some semantic coherence is noted and that the roots act like a variable within these structures. As such, these formulate semantic structures with actual spaces left for the root. One may compare the VerbNet patterns for *eat* and *feed*. We see that by characterizing the patterns for generative derivations, the meaning of large number of terms may be generated:

(3) Pattern $C^1aC^2aC^3a$.1:

root (during(E), Agent, Patient)

+root Akl: take_in

   = >akal: take_in (during(E), Agent, Patient)

+root lqm: take_in

   = Ailotaqam: take_in(during(E), Agent, Patient)

compare to English: eat-39.1: take_in(during(E), Agent, Patient)

(4) Pattern $aC^1aC^2aC^3a$:

root(during(E), Recipient, Patient) cause(Agent, E)

+root Akl: take_in

   = >akal: take_in(during(E), Agent, Patient) cause(Agent, E)

+root lqm: take_in

   = Ailotaqam: take_in(during(E), Agent, Patient) cause(Agent, E)

compare to English: feeding-39.7: take_in(during(E), Recipient, Patient)

cause (Agent, E)

As such, one may see that while this pattern lexicon does not form the same sort of semantically coherent verb classifications seen in a VerbNet-style approach, one could easily collect these syntactic patterns with semantically similar roots to form semantically coherent verb classes. How this would compare with other automated verb classifications (such as in [20]) remains to be seen. These entries would also be developed for the "deverbal" morphemes, such as the verbal nouns, active participles, etc., of each sense. These deverbal morphemes would mostly be automatically derivable from their originating patterns, and would otherwise be done by hand in a fashion similar to the above methods.

D. Annotating Patterns of Derivation

Finally, a third set of tasks would involve marking the derivational linkages between these various entries, both semantic and syntactic. This would involve going through each lexical entry described in Section 4.3, gathering data concerning its presumed derivations (subtle shifts in semantics being a core marker of this phenomenon), and marking direct linkages to other senses in other patterns, based upon the data at hand and existing works [2].

For example, while it is well known that pattern V, taCaCCaCa, is often derived from pattern II, CaCCaCa, it really only derives from a subset of senses within pattern II. This would contribute additional value to the larger task of identifying the sense of a given morpheme pattern, as it allows one to use the lexicon in order to obtain more accurate disambiguation of novel verbal instances.

This linkage would be applied to the deverbal forms as well. This would provide mappings for how deverbal forms develop, while also indicating which senses are more productive for various deverbal forms. This would not only aid in the task of disambiguating new words, but could also aid in simpler tasks of morphological analysis, adding a more nuanced statistical measure of probability to derivation rules which are currently represented only as possible or impossible.

These linkages would not simply mark one term as being derived from another, but would also map semantic role labels onto each other. As such, marking the derivation which starts with the reciprocal sense of [C1aaC2aC3a] (having transitive syntax for phrases like 'John quarreled with Mary') and results in the reflexive [taC1aaC2aC3a] pattern (with intransitive meanings like 'John and Mary quarreled'), this sort of coding would occur in the reflexive [taC1aaC2aC3a] sense:

[C1aaC2aC3a]. reciprocal roles

Agent   Patient

[taC1aaC2aC3a]

Agent {+plural}
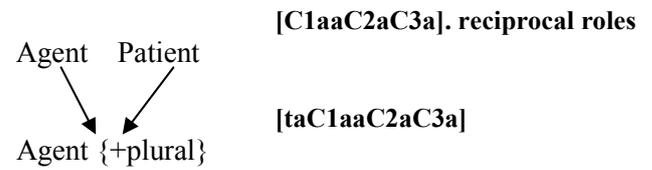
The full lexical entry is as follows:
iC1otaC2aC3a-1

Prose description: reflexive

Roles: Arg0:   Agent<+int_control>=Theme**,**   Arg1: location

Syntax: V Agent=Theme**;** V Agent=Theme {prep} Location

Semantics: ROOT (during(E), Agent, ?Theme)

Derived   from   senses: [$C^1aC^2aC^3a$-ya$C^1aC^2iC^3$][$C^1aC^2aC^3a$-ya$C^1aC^2uC^3$]   [$C^1aC^2aC^3a$-ya$C^1aC^2aC^3$]

Table 2 illustrates pattern verb senses. These patterns look longer because a set of the simple verb paradigms are the same in the perfective form and require marking of the imperfective to distinguish them.

TABLE II.   EXAMPLE VERB-SENSES

AixotafaY	اختفى	To disappear (from s.t.)
Aigotasal	اغتسل	To bathe o.s
AibotaEad	ابتعد	To refrain (from s.t.)
AimotanaE	امتنع	To content o.s (with s.t.)
AikotafaY	اكتفى	To exhaust o.s (with s.t.)
Aijotahad	اجتهد	To retire/withdraw o.s (from s.t.)
Aixotaba	اختبأ	To hide (away from s.t)

Finally, Table 3 shows the results of our study of 10 morphological patterns through the annotation of the proposed AMPN verb frames for all the verbs applied to the examined patterns and included in the APB. The first column represents the three levels of the assumed derivational hierarchy, the second column represents the morphological

patterns and number of verb frames which occurred in APB file, the third column represents the verb pattern frames classified according to their syntactic alternations, and the fourth column includes the number of APB verb frames which apply to each verb pattern frame.

TABLE III. SUMMARY RESULT OF THE PILOT STUDY OF THE AMPN

Level	MP (# of frames in APB)	Verb Pattern Frame	Number of APB Verbs
2	استَفعَل AisotaC1oC2aC3 (179)	F_1: Causative Verbs	98
		F_2: Ergative Verbs	96
		F_3: Inchoative Verbs	52
		F_4: Reflexive Verbs	23
2	إنْفَعَل AinoC1aC2C3 (111)	F_1: Causative Verbs	6
		F_2: Ergative Verbs	33
		F_3: Inchoative Verbs	21
		F_4: Reflexive Verbs	61
2	إفْتَعَل AiC1otaC2aC3 (342)	F_1: Causative Verbs	131
		F_2: Ergative Verbs	40
		F_3: Inchoative Verbs	42
		F_4: Reciprocal Verbs	16
		F_5: Reflexive Verbs	113
2	أفعَل >aC1oC2aC3 (289)	F_1: Causative Verbs	265
		F_2: Inchoative Verbs	15
		F_3: Reflexive Verbs	9
2	فَاعَل C1AC2aC3 (146)	F_1: Causative Verbs	59
		F_2: Inchoative Verbs	4
		F_4: Reciprocal Verbs	82
3	تَفَاعَل taC1AC2aC3 (125)	F_1: Causative Verbs	18
		F_2: Ergative Verbs	5
		F_3: Inchoative Verbs	8
		F_4: Reciprocal Verbs	82
		F_5: Reflexive Verbs	10
2	فَعَّل C1aC2C2aC3 (357)	F_1: Causative Verbs	352
		F_4: Reciprocal Verbs	2
		F_4: Reflexive Verbs	3
3	تَفَعَّل taC1aC2C2aC3 (225)	F_1: Causative Verbs	26
		F_2: Ergative Verbs	64
		F_3: Inchoative Verbs	68
		F_4: Reciprocal Verbs	33
		F_5: Reflexive Verbs	64
1	فَعْلَل C1aC2C3aC4 (22)	F_1: Causative Verbs	15
		F_2: Inchoative Verbs	1
		F_3: Reciprocal Verbs	1
		F_4: Reflexive Verbs	5
2	تَفَعْلَل taC1aC2C3aC4 (8)	F_2: Ergative Verbs	3
		F_4: Reflexive Verbs	5

V. CONCLUSION AND FUTURE WORK

The present paper introduced a pilot study for a Lexical Semantic Resource for Arabic Morphological Patterns. The goal of this Project was to represent syntactic valancy and Alternation of Arabic Morphological Pattern. The current pilot work was built on Arabic PropBank for Arabic verb Lexicon (2100 Verb Frame). We annotated and classified APB verbs based on Argument Structure and aspectual features. The result of analysis represents the possible frames for each pattern which could be considered an infrastructure for an Arabic Lexicon. The future work is to build a full-sized resource for Arabic Morphological Pattern which covers verb and derivational lexicon, to build a comprehensive lexical semantic resource for AMPN, and to map it to several lexical resources such as Arabic PropBank, Arabic WordNet and Arabic VerbNet.

REFERENCES

[1] Abdul-Raof Hussein. 1997. Rhetorical word order in Arabic in the light of modern European linguistic theories'. *In Journal of Language Studies*, volume 9, (Khartoum International Institute of Arabic Language, Arab League Educational, Cultural and Scientific Organization) 1997, pp.35-46.

[2] Al-Najem, S. R. 2007. Inheritance-based approach to arabic verbal Root-and-Pattern morphology. *Text, Speech and Language Technology* 38 (II), 67–88.

[3] Almutawakkal, A. 1988. *qadaya mu'agima: almuhamulat alfi'aliya almustaqa fi allaga al arabiy*. Rabaat, Morocco: athaad alnaaSriin almagaaribiya.

[4] Hayder K. Al Ameed, Shaikha O. Al Ketbi, Amna A. Al Kaabi, Khadija S. Al Shebli, Naila F. Al Shamsi, Noura H. Al Nuaimi, Shaikha S. Al Muhairi. 2005. Arabic light stemmer: A new enhanced approach. *In 2nd International Conference on Innovations in Information Technology (IIT'05)*, Dubai, United Arab Emirates.

[5] Attia Mohammed, Pavel Pecina, Lamia Tounsi, Antonio Toral, Josef van Genabith. 2011. An Open-Source Finite State Morphological Transducer for Modern Standard Arabic. International Workshop on Finite State Methods and Natural

[6] Attia Mohammed .2006. An Ambiguity-Controlled Morphological Analyzer for Modern Standard Arabic Modelling Finite State Networks. *The Challenge of Arabic for NLP/MT Conference*, October 2006. The British Computer Society, London.

[7] Beesley, K. R. 2001. Finite-state morphological analysis and generation of arabic at xerox research: Status and plans in 2001. In *ACL Workshop on Arabic Language Processing: Status and Perspective*, Volume 1, pp. 1–8. Citeseer.

[8] Benmamoun, E. 2003. The Role of the Imperfective Template in Arabic Morphology. Chapter Book. Ed. Joseph Shimron. John Benjamins, 2003.

[9] Buckwalter, T. 2004. Buckwalter Arabic Morphological Analyzer Version 2.0. LDC Catalog No: LDC2004L02, Linguistic Data Consortium, Philadelphia.

[10] Diab, Mona, Alessandro Moschitti, Daniele Pighin. 2008. Arabic Semantic Role Labeling. In *Proceedings of the Conference of American Association for Computational Linguistics (ACL'08)*.

[11] Diab, Mona. 2009. Second Generation Tools (AMIRA 2.0): Fast and Robust Tokenization, POS tagging, and Base Phrase Chunking. In *MEDAR 2nd International Conference on Arabic Language Resources and Tools,* April, Cairo, Egypt.

[12] Farghaly, A. and J. Senellart .2003. Intuitive coding of the arabic lexicon. In SYSTRAN, MT, Summit IX Workshop, *Machine Translation for Semitic Languages: Issues and Approaches*, Tuesday September, Volume 23.

[13] Fehri, A. F. 1988. *albinaa' almuwaazii nazriya fii bnaa' alkalamaa wa binaa' aljumlaa*. Altabi'aa raqum 2 daar tuwabqaal lilnaSar aldaar albaydaa'.

[14] Habash, Nizar and Owen Rambow. 2005. Arabic Tokenization, Morphological Analysis, and Part-of-Speech Tagging in One Fell Swoop. In *Proceedings of the Conference of American Association for Computational Linguistics (ACL'05)*.

[15] Kipper, K., A. Korhonen, N. Ryant, and M. Palmer 2008. A large-scale classification of english verbs. *Language Resources and Evaluation Journal* 42(1), 21–40.

[16] Maamouri, M., Dave Graff, Basma Bouziri, Sondos Krouna, Ann Bies, Seth Kulick .2010. LDC Standard Arabic Morphological Analyzer (SAMA) Version 3.1. LDC Catalog No.: LDC2010L01, Linguistic Data Consortium, Philadelphia.

[17] Marton, Y., N. Habash, and O. Rambow .2010. Improving arabic dependency parsing with lexical and inflectional morphological features. In *The First Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL 2010)*, pp. 13.

[18] Mousser, Jaouad .2011. Classifying arabic verbs using sibling classes. In *the Proceedings of the International Workshop on Computational Semantics*, Oxford, UK.

[19] Palmer, M., O. Babko-Malaya, A. Bies, M. Diab, M. Maamouri, A. Mansouri, and W. Zaghouani .2008. A pilot arabic propbank. In Proceedings of LREC 2008, pp. 28–30.

[20] Snider, N. and M. Diab .2006. Unsupervised induction of modern standard arabic verb classes. In *Proceedings of the Human Language Technology Conference of the NAACL*, Companion Volume: Short Papers on XX, pp. 153–156. Association for Computational Linguistics.

[21] Abdelhadi Soudi, Antal van den Bosch and Günter Neumann. 2007. Arabic Computational Morphology Knowledge-based and Empirical Methods. *Text, Speech and Language Technology*.Volume 38, 2007.

[22] The Challenge of Arabic for NLP/MT Conference .2006. An Ambiguity-Controlled Morphological Analyzer for Modern Standard Arabic Modelling Finite State Networks. The Challenge of Arabic for NLP/MT Conference: The British Computer Society.

[23] Watson Janet C. E. 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press. Pp. v+307.

[24] Zaghouani, W., M. Diab, A. Mansouri, S. Pradhan, and M. Palmer .2010. The revised arabic PropBank. In Proceedings of the Fourth Linguistic Annotation Workshop, Uppsala, Sweden, pp. 222–226.