

Classifying Arabic Verbs Using Sibling Classes

Jaouad Mousser
University Of Konstanz
Department of Linguistics
Jaouad.Mousser@uni-konstanz.de

Abstract

In the effort of building a verb lexicon classifying the most used verbs in Arabic and providing information about their syntax and semantics (Mousser, 2010), the problem of classes over-generation arises because of the overt morphology of Arabic, which codes not only agreement and inflection relations but also semantic information related to thematic arity or other semantic information like "intensity", "pretension", etc. The hierarchical structure of verb classes and the inheritance relation between their subparts expels derived verbs from the main class, although they share most of its properties. In this article we present a way to adapt the verb class approach to a language with a productive (verb) morphology by introducing sibling classes.

1 Introduction

Class based approach to lexical semantics such as presented in Levin (1993) provides a straightforward way of describing a large number of verbs in a compact and generalized way. The main assumption is the correlation between the syntactic behaviour of verbs as reflected in diathesis alternations and their semantic properties. Verbs which participate in the same set of diathesis alternations are assumed to share the same meaning facets. Verbs like *abate*, *acidify*, *dry*, *crystallize*, etc. share a meaning component and are grouped into a class (change-of-state), since they participate in the *causative/incoative alternation*, the *middle alternation*, the *instrument subject alternation* and the *resultative alternation* (Levin, 1993). Class based lexica have turned out to be useful lexical resources such as the English VerbNet (Kipper Schuler, 2005), which provides information about thematic roles, syntactic and semantic structure of 5879 English verbs. Trying to use the same approach to classify verbs of a morphologically rich language like Arabic, the researcher is faced with difficulties because many alternations require morphological operations to express meaning aspects, especially those related to thematic roles.

(1) Causative/Incoative Alternation in Arabic

- a. *naššafa saliymun ālmalābisa.*
dry-CAUS-PRF Salim-SUBJ-NOM DEF-cloth-PL-OBJ-ACC.
'Salim dried the clothes.'
- b. *našafati ālmalābisu.*
dry-PRF-PL DEF-cloth-PL-SUBJ-NOM
'The colthes dried.'

In example (1) the causative/incoative alternation is realized through an overt morphological change on the head of the sentence (reduplication of the second root consonant in (1a)), in such a way that the verb changes to a new entry, which according to the hierarchical organisation of the class and especially to the inheritance relation between its subparts, cannot longer be kept into the original class. Transporting the new verb entry into a new class risks to loose its connection to the original class, which is an undesired effect, since it does not necessarily reflect the natural organisation of the lexicon of Arabic.

2 Arabic VerbNet and Class Structure

Arabic VerbNet¹ is a large coverage verb lexicon exploiting Levin's classes (Levin, 1993) and the basic development procedure of Kipper Schuler (2005). The current version has 202 classes populating 4707 verbs and 834 frames. Every class is a hierarchical structure providing syntactic and semantic information about verbs and percolating them to subclasses. In the top level of each class there are verb entries represented as tuples. Each tuple contains the verb itself, its root form, the deverbal form and the participle. At the same level thematic roles and their restrictions are encoded. The important information about the class resides in the frames reflecting alternations where the verbs can appear. Every frame is represented as an example sentence, a syntactic structure and a semantic structure containing semantic predicates and their arguments and temporal information in a way similar to Moens and Steedman (1988). Every class can have subclasses for cases where members deviate from the prototypical verb in some non central points. A subclass recursively reflects the same structure as the main class and can (therefore) itself have subclasses. A subclass inherits all properties of the main class and is placed in such a way that the members in the top level are closed for the information it adds. This fact hinders putting derived verbs participating in alternations into the main class or in one of the subclasses.

3 Sibling Classes

Introducing sibling classes is a way to resolve the problem arising from the discrepancy between two derivationally related morphological verb forms which participate in the same set of alternations and therefore share the same semantic meaning. Tables 1 and 2 show two sibling classes and their alternations sets. The incoative alternation introduces a morphological change in the verbs. This fact blocks the derived verbs from entering in any inheritance relation to the base verbs according to the hierarchical structure of the class they belong to. Consequently, a sibling class (Table 2) is created to populate the verbs resulting from alternations requiring morphological changes.

4 Automatic Extension of Arabic VerbNet via Sibling Classes

4.1 Morphological Verb Analyser

In order to generate derived verb forms a Java based morphological analyser was implemented as part of a system in order to generating sibling classes automatically (Sibling class generator SCG). This provides an analyse of the morphological composition of the input verbs. The program is based on regular expressions and identifies the following features:

- *Verb root*: This corresponds to an abstract form of 2–4 consonants carrying a basic semantic meaning of the verb. Thus, *ktb* is the abstract root of the verb *kataba* 'to write' but also of other derivationally related words such as *linkataba* 'INC-write', *takaAtaba*, 'RECIP-write' 'to correspond'.
- *Verb pattern*: This corresponds to the verb pattern in the classical Arabic grammar and is represented by a canonical verb form *faEala*² where the letters *f*, *E* and *l* correspond respectively to the first, the second and the third root consonant of the input verb. Thus, the pattern of a verb such as *Iinokataba* will be *IinofaEala*, where *f*, *E* and *l* correspond to *k*, *t*, *b* which are the root consonants of the verb.

Table 3 shows the produced morphological analysis of the verbs *kataba* 'to write', *Iinokataba* 'INC-write' and *takaAtaba* 'to correspond'. The extracted features are then used in combination with semantic information of verb classes to generate morpho-semantic derivational forms of verbs and later semantically derived verb classes (sibling classes) as explained in the next sections.

4.2 Identifying Expandable Verb Classes

The input of SCG are the basic verb classes produced in the first stadium of the lexicon building (Mousser, 2010). In order to define which classes are good candidates to be expanded according to

¹http://ling.uni-konstanz.de/pages/home/mousser/files/Arabic_VerbNet.php

²Pattern are transliterated using Buckwalter's style. All other Arabic examples are transliterated using Lagally

Table 1: The *change of state* class in Arabic. The causative use.

Class: Change of State			
Members: <i>aṣṣarana</i> ‘modernize’, <i>ḥaṣḥaṣa</i> ‘privatize’, <i>awolama</i> ‘globalize’, <i>arraba</i> ‘arabize’, etc.			
Roles and Restrictions: Agent [+int_control] Patient Instrument			
Descriptions	Examples	Syntax	Semantics
Basic Intransitive	<i>naṣṣafa salīm malābisahu.</i> (Salim dried his clothes)	V Agent Patient	<i>cause(Agent, E), state(result(E), Endstate, Patient)</i>
NP-PP	<i>naṣṣafa salīm malaābisahu biālbūḥāār.</i> (Salim dried his clothes with the vapour)	V Agent Patient {bi} Instrument	<i>cause(Agent, E), state(result(E), Endstate, Patient), use(during(E), Agent, Instrument)</i>
Instrument Subject	<i>naṣṣafa ālbūḥāāru ālmalābisa.</i> (The vapour dried the clothes.)	V Instrument Patient	<i>use(during(E), ?Agent, Instrument), state(result(E), Endstate, Patient)</i>
Subclass			

Table 2: The *change of state* sibling class in Arabic. The incoative use.

Sibling Class: Change of State			
Members: <i>taṣṣarana</i> ‘INC-modernize’, <i>taḥḥaṣa</i> ‘INC-privatize’, <i>taawolama</i> ‘INC-globalize’, <i>taarraba</i> ‘INC-arabize’, etc.			
Roles and Restrictions: Agent [+int_control] Patient Instrument			
Descriptions	Examples	Syntax	Semantics
V NP.patient	<i>naṣṣafati ālmalābisahu.</i> (The clothes dried)	V Patient	<i>state(result(E), Endstate, Patient)</i>
PP	<i>naṣṣafati ālmalābisahu biālbūḥāār.</i> (The clothes dried with the vapour.)	V Patient Instrument	<i>use(during(E), ?Agent, Instrument), state(result(E), Endstate, Patient)</i>
Subclass			

causativity criteria, thematic role information and semantic predicates of class frames are detected. Classes of verbs with the thematic role *agent* and compositional semantics containing the causative predicate *CAUSE* are selected as in the case of *change-of-state* classes. Additionally, inherently uncausative verb classes involving a change of state are identified according to whether they possess a *patient* theme occupying the subject position and accordingly whether their compositional semantics include the change of state predicate *STATE*.

4.3 Generating Sibling Classes

Generating sibling classes requires generating the appropriate morphological verb forms, new lists of thematic roles and new frames with new syntactic descriptions and new predicate semantics reflecting the derived meaning of the verbs (See Tables 1 and 2).

4.3.1 Generating New Verb Forms

Verbs of the new sibling classes are generated from morphological forms of the base verbs using the following information:

- The semantic morphological operation required for the input class (causativization, reciprocalization or decausativization).
- The morphological properties of the input verbs such as root, pattern and segmental material.
- Rewrite rules defining for each input verb pattern the appropriate derivative form to express the target semantic meaning.

The generation of derived verbs reveals itself to be the reverse of the morphological analysis, as it consists of replacing the consonants *f*, *E* and *l* of the relevant output pattern with the root consonants of the input verb. Thus, the *change-of-state* verb *faḥḥama* ‘to carbonize’ with the root *fḥm* and the pattern faEāla will produce the derived verb *tafaḥḥama* ‘INC-carbonize’ according to the decausativization rule 2 in the Table 4 and by replacing the output pattern consonants *f*, *E* and *l* respectively with the root consonants *f*, *ḥ* and *m*.

Table 3: Morphological information

Verb	Root	Pattern	Segments
<i>kataba</i>	ktb	faEala	.a.a.a
<i>Iinokataba</i>	ktb	IinofaEala	Iino.a.a.a
<i>takaAtaba</i>	ktb	taFaAEala	ta.aA.a.a

Table 4: Rewrite rules for decausativization

Input pattern	Output pattern
faEala	⇒ IinofaEala
faEāla	⇒ tafaEāla
faAEala	⇒ tafaAEala
faEolana	⇒ tafaEolana
fawoEala	⇒ tafawoEala

4.3.2 Generating New Lists of Thematic Roles

Building sibling classes is not only a morphological process but also a semantic one with repercussions on the thematic arity of the concerned class. Thus, the simple reciprocal alternation found with social interaction and communication verbs adds a new theme role *actor* which can be used interchangeably with the two symmetrical themes *actor1* and *actor2*. Other operations delete thematic roles in the new class. Thus decausativization deletes the thematic role agent from the list of roles.

4.3.3 Generating New Argument Structures

Adapting thematic structures of the new sibling classes has an influence on their argument structures. Thus, adding a new thematic role while causativizing a verb class is reflected in the syntactic level by adding a new argument with its appropriate restrictions. For instance, the introduction of the theme *actor* in the simple reciprocal alternation of interaction verbs imposes an additional restriction [+*dual*/+*plural*] on the subject at the syntactic level, whereas the object is omitted from the argument structure of the concerned frame. Additionally, the mapping between thematic roles and grammatical arguments is the subject of change. Thus, change-of-state verbs and other causative verbs are reflexivized by assigning a *agent* role to the *patient* in the causative reading. At the syntactic level this operation is reflected by omitting the subject and promoting the object to the subject position.

4.3.4 Generating New Semantic Descriptions

For sibling classes to reflect the meaning variations introduced by the new morphological material, the semantic description of input classes has to be modified by adding or omitting appropriate semantic predicates. Thus, causativization introduces the predicate *CAUSE* to the semantic description of the class, whereas decausativization is reflected by omitting the same predicate and its argument which corresponds mostly to the *agent* of the concerned frame. In the case of a simple reciprocal alternation the presence of one (plural) *actor* is reflected by introducing two presupposed (implicit) *actor* roles: *actor_i* and *actor_j* in the main semantic description of the verb as shown in (2) in contrast to explicit *actor* roles in (3).

(2) **Implicit symmetrical actor roles**

social_interaction(during(E), Actor_i, Actor_j)

(3) **Explicit symmetrical actor roles**

social_interaction(during(E), Actor1, Actor2)

4.3.5 Generating New Frames

We generate new frames (alternations) on the basis of frames of the base (input) classes. Since operations like decausativization affect only the thematic arity of the class, alternations which are not related to causativity are reproduced in the new classes. For instance, the frame for the instrumental alternation of the causative verb class is reproduced by adapting the thematic structure to the incoative use. Thus,

the frame alternation of (4a) will produce the frame alternation (4b), since the instrumental alternation in Arabic can be found with causative verbs as well as with uncausative verbs.

(4)

- a. *naššafa saliymun ālmalābisa. biālbuhaāri*
 dry-CAUS-PRF Salim-SUBJ-NOM DEF-cloth-PL-OBJ-ACC with-DEF-vapor.
 ‘Salim dried the clothes with the vapor.’
- b. *našifati ālmalābisu. biālbuhaāri*
 dry-PRF DEF-cloth-PL-SUBJ-NOM with-DEF-vapor.
 ‘The clothes was dried with the vapor.’

5 Results and Discussion

We run SCG on the current version of Arabic VerbNet. The program was able to identify 89 expandable classes with 3005 verbs and 368 frames, 60 of them populate causative and 29 uncausative verbs. For each class one sibling class was generated with a total of 3360 verbs and 368 frames. The high number of generated verbs is due to the fact that some verbs have more than one way to express the causative or the inchoative. After checking the quality of the produced classes, we count 71% accuracy in identifying the patterns of the verbs and 82% in generating their derived forms. After manually adjusting the new sibling classes (deleting unsuitable verb forms and adding the correct ones, adding frame examples, etc.), we noted that Arabic VerbNet counts now 291 classes populating 7937 verbs and 1202 frames, which represents an expansion rate of 44%. Noteworthy, not all verbs formed by the root-pattern system exist synchronically. We observed that inside the same sibling class one verb can be widely found in different Arabic corpora whereas another verb of the same sibling class is not attested in the same corpora. For instance, the verb *nabaḥa* ‘to bark’ of the class *animal_sounds* has a causative form *anbaḥa* ‘cause to bark’, but for the most members of the same class the causative form are not attested to be used in the “real world”. However, they are potential lexicon entries and native Arabic speakers will most likely recognize their meaning without being exposed to them before. Additionally, given the fact that human lexica are brittle and incomplete, the scope of Levin’s class approach (Levin, 1993) can be expanded to explain the derivational behaviour of verbs: Verbs which belong to the same class and share the same syntactic and semantic properties are likely to share the same derivational behaviour, especially when this behaviour is related to the general semantic properties of the class.

6 Conclusion

We presented a way to classify verbs of a language with a productive (verb) morphology like Arabic. Additionally to the traditional classes with a rigid hierarchical structure and a top-down inheritance relation, sibling classes were introduced to classify those verbs which engage in morphological operations during diathesis alternations. Sibling classes are autonomous classes which maintain relations to the class they are issued from consequently reflecting the natural connection between parents element in the lexicon.

References

- Kipper Schuler, K. (2005). *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph. D. thesis, University of Pennsylvania.
- Korhonen, A. and T. Briscoe (2004). Extended lexical-semantic classification of english verbs. In *The HLT/NACCL wokshop on computational lexical semantics*.
- Levin, B. (1993). *English Verb Classes and Alternations. A Preliminary Investigation*. Chicago and London: The University of Chicago Press.
- Moens, M. and M. Steedman (1988). Temporal ontology and temporal reference. *Computational Linguistics 14*, 15–28.
- Mousser, J. (2010). A large coverage verb taxonomy for arabic. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC’10)*, Valetta, Malta.