5

Inheritance-based Approach to Arabic Verbal Root-and-Pattern Morphology

Salah R. Al-Najem

Department of Arabic, Faculty of Arts, Kuwait University, P.O. Box: 23558, Safat, Kuwait
alnajem@arts.kuniv.edu.kw

Abstract: This chapter introduces a computational approach to the derivation and inflection of Arabic verbs. The approach attempts to capture generalizations, dependencies, and syncretisms existing in Arabic verbal morphology in a compact non-redundant manner. The approach is represented by a computational implementation in DATR lexical knowledge representation language. Generalizations will be captured through the use of Default Inheritance technique. Dependencies will be handled through the use of Multiple Inheritance technique. Default Inference technique will be used to capture syncretisms

5.1 Introduction

Arabic morphology is a system governed by a number of generalizations, dependencies, and syncretisms that explain the overt aspects of regularity in the derivational and inflectional structure of Arabic. A generalization is a statement about the facts of a language, which holds true in all cases or in nearly all cases (Trask 1993). A dependency is a case in which a (dependent) form is derived from another form. Finally, a syncretism is the case in which two or more morphemes that are morphosyntactically distinct appear identical in form.

In the context of implementing Arabic morphology computationally, these generalizations, dependencies, and syncretisms should be taken into consideration. Capturing such generalizations, dependencies, and syncretisms in a computational implementation of Arabic morphology can save that implementation from redundancy and can make it more compact, which is a vital issue in linguistics and computer science. This chapter introduces a computational approach to Arabic verbal morphology in which these generalizations, dependencies, and syncretisms are used to systematize Arabic in a concise manner. The approach is represented by an implementation written in the DATR lexical knowledge representation language.

A. Soudi, A. van den Bosch and G. Neumann (eds.), Arabic Computational Morphology, 67–88.

5.2 Linguistic Data

Verbs in Arabic are formed from roots consisting of three or four letters (known as radical letters). From these roots, verbal stems are constructed using a number of canonical forms known as measures. Measures are sequences of consonants and vowels that represent word structure. They may also contain stem derivational affixes. Each measure is normally associated with perfective (active and passive), imperfective (active and passive), and imperative patterns, which are used to form perfective, imperfective, and imperative verbal stems. The perfective verbs indicate a completed act, while imperfective verbs denote an unfinished act, which is just beginning or in progress. Measures that are intransitive or express a state of being do not normally associate with passive voice inflection. The stems formed using the above patterns are used to construct verbs through prefixing and/or suffixing inflectional prefixes and/or suffixes.

Verbs in Arabic are either triliteral (having three radical letters) or quadriliteral (having four). The triliteral verbal stems are formed using fifteen verbal measures while the quadriliteral verbal stems use four. Table 5.1 shows seven of these measures with examples that demonstrate how verbal stems are constructed using them. The stems are given in this table without inflectional prefixes and suffixes. For the triliteral verbs, the root *ksr* (breaking) will be used as an example root, while the root *dHrj* (rolling) will be used for the quadriliteral verbs. Roman numbers are used for reference with the prefix Q denoting a quadriliteral verbal measure. It should be

Table 5.1. The verbal measures

No	Measure	Active Perfective	Passive Perfective	Active Imperfective and Imperative	Passive Imperfective
I	faς al	$C_1aC_2aC_3$ kasar	$C_1uC_2iC_3$ kusir	$C_1C_2iC_3$ ksir	$C_1C_2aC_3$ ksar
II	faς~al	$C_1aC_2{\sim}aC_3$ kas~ar	$C_1uC_2{\sim}iC_3$ kus~ir	$C_1aC_2{\sim}iC_3$ kas~ir	$C_1aC_2{\sim}aC_3$ kas~ar
III	faAςal	$C_1aAC_2aC_3$ kaAsar	$C_1uwC_2iC_3$ kuwsir	$C_1aAC_2iC_3$ kaAsir	$C_1aAC_2aC_3$ kaAsar
IV	Âafςal	$\hat{A}aC_1C_2aC_3$ Âaksar	$\hat{A}uC_1C_2iC_3$ Âuksir	$\hat{A}aC_1C_2iC_3$ Âaksir	$\hat{A}aC_1C_2aC_3$ Âaksar
V	tafaς~al	$taC_1aC_2{\sim}aC_3$ takas~ar	$tuC_1uC_2{\sim}iC_3$ tukus~ir	$taC_1aC_2{\sim}aC_3$ takas~ar	$taC_1aC_2{\sim}aC_3$ takas~ar
QI	faςlal	$C_1aC_2C_3aC_4$ daHraj	$C_1uC_2C_3iC_4$ duHrij	$C_1aC_2C_3iC_4$ daHrij	$C_1aC_2C_3aC_4$ daHraj
QII	tafaςlal	$taC_1aC_2C_3aC_4$ tadaHraj	$tuC_1uC_2C_3iC_4$ tuduHrij	$taC_1aC_2C_3aC_4$ tadaHraj	$taC_1aC_2C_3aC_4$ tadaHraj

noted that not all measures are applicable with every root. The roots *ksr* and *dHrj*, for example, do not occur with some measures. For instance, **Âaksar* (Measure IV) is unacceptable in Standard Arabic while *kasar* (Measure I) is acceptable. It should also be noted that some stems undergo certain additional phonological processes. In addition, in this table, I will use CV arrays to represent the perfective, imperfective, and imperative stems (patterns) of each measure.[1]

Measure I (the stem *kasar*) is considered the basic measure (stem) from which the other 14 triliteral verbal stems are derived. The derivation is done throgh various modifications of the meaning associated with that basic stem. Similarly, measure QI (the stem *daHraj*) can be considered a basic form (stem) from which the remaining three quadriliteral verbal stems are derived.

So far, we have considered the derivation of verbs. We turn now to verb inflection. The inflection of verbs in Arabic is mainly achieved through the use of prefixes and suffixes denoting person, number, and gender. These non-stem prefixes and suffixes are affixed to the perfective, imperfective, and imperative stems, which are produced using the verbal measures from roots as shown in Table 5.1. The reader should note that further changes in stem structure, such as radical letter rejection, vowel rejection, and radical letter transformation, are also applied in the inflection process when we deal with some roots.

As mentioned above, Arabic verbs have perfective, imperfective, and imperative stems. The perfective verbs indicate a completed act, while imperfective verbs denote an unfinished act which is just commencing or in progress. Table 5.2 shows a partial inflection paradigm illustrating the perfective active, imperfective active, and imperative inflection for the second person masculine and feminine.

Table 5.2. The perfective, imperfective, and imperative paradigm

Perfective		Imperfective	Imperative
2nd Person Singular			
Mas	kasar-ta	ta-ksir-u	ksir[2]
Fem	kasar-ti	ta-ksir-iyn	ksir-iy
2nd Person Dual			
Mas	kasar-tumaA	ta-ksir-aAn	ksir-aA
Fem	kasar-tumaA	ta-ksir-aAn	ksir-aA
2nd Person Plural			
Mas	kasar-tum	ta-ksir-uwn	ksir-uw
Fem	kasar-tun~a	ta-ksir-na	ksir-na

[1] So a measure can actually produce up to four different stems.

[2] The imperfective stem *ksir* has the surface form *Ǎiksir* with the epenthetic *Ǎi* prefixed to it.

There are a number of generalizations that can be captured in relation to Arabic verbal inflections. These generalizations are:

(1) Generalizations about Verbal Inflectional Paradigms:

a) Perfective active and passive inflectional paradigms are constructed by suffixing a perfective inflectional suffix (like: {-*tu*}) to a perfective stem (like *kasar*). There are no perfective inflectional prefixes.
b) Imperfective active and passive inflectional paradigms are constructed by prefixing an imperfective inflectional prefix (like:{*ya*-}) to an imperfective stem (like: *staksir*) and suffixing an imperfective inflectional suffix (like {-*u*}).
c) Imperative inflectional paradigms are constructed by suffixing an imperative inflectional suffix (like: {-*aA*}) to an imperative stem (like: *staksir*). There are no imperative inflectional prefixes.
d) The imperative stems are the same as the imperfective active stems.
e) The imperfective active prefixes end with the vowel '*a*.' The imperfective passive prefixes are the same as the imperfective active prefixes except that the prefix vowel becomes '*u*.'

Besides the previous generalizations relating to the formation of verbal paradigms, there is a sub-generalization related to the inflectional paradigms that correspond to the verbal measures II, III, IV, and QI. The inflectional paradigms that correspond to these measures are constructed in the same way that the verbal inflectional paradigms mentioned in (1) are except that the imperfective active prefix will end with the vowel '*u*' instead of '*a*.' In other words, the vowel '*a*' of the imperfective active prefixes is changed to '*u*' when these prefixes are used with the imperfective active stems that consist of exactly two heavy syllables. A heavy syllable is a syllable with a branching nucleus (CVV) or a branching rime (CVC). These dual heavy syllable stems are those imperfective active stems produced using the measures, II, III, IV, and QI, such as the stem *kas~ir* (Measure II).

In addition, there are dependencies between verbal measures. There are, in fact, overt dependencies between the basic verbal measure I and most of the derived triliteral verbal measures. There are also overt dependencies between the basic quadriliteral verbal measure QI and most of the derived quadriliteral verbal measures. We can capture these dependencies by means of the following generalizations:

(2) Dependencies between Verbal Measures:

a) The perfective active and passive patterns (stems) of the triliteral derived measures are constructed with the perfective active and passive patterns (stems) of the basic triliteral measure I ($C_1aC_2aC_3$[3] and $C_1uC_2iC_3$) by

[3] The final syllable vowel (*a*) of Measure I perfective active pattern ($C_1aC_2\mathbf{a}C_3$) is changed to '*u*' or '*i*' according to stem root (see Table 5.1). For example, the perfective active stem formed from the roots *Hsn* and *ςlm* using Measure I are *Has**u**n* ($C_1aC_2\mathbf{u}C_3$) and *ςal**i**m* ($C_1aC_2\mathbf{i}C_3$) not * *Has**a**n* and * *ςal**a**m*.

changing the pattern (stem) portion that precedes the final C_2VC_3 syllable of these derived patterns (stems).
The imperfective active and passive patterns of the triliteral derived measures are constructed with the imperfective active and passive patterns of the basic triliteral measure I ($C_1C_2iC_3$[4] and $C_1C_2aC_3$) by changing the pattern portion that precedes the final C_2VC_3 syllable of these derived patterns.

b) The perfective active and passive patterns of the quadriliteral derived measures are derived from the perfective active and passive patterns of the basic quadriliteral measure QI ($C_1aC_2C_3aC_4$ and $C_1uC_2C_3iC_4$) by changing the pattern portion that precedes the final C_3VC_4 syllable of these derived patterns.
The imperfective active and passive patterns of the quadriliteral derived measures are derived from the imperfective active and passive patterns of the basic quadriliteral measure QI ($C_1aC_2C_3iC_4$ and $C_1aC_2C_3aC_4$) by changing the pattern portion that precedes the final C_3VC_4 syllable of these derived patterns.
The changes include the gemination of radical letters, the deletion of vowels, and the insertion of affixes. For example, the Measure II perfective active pattern C_1aC_2 ~aC_3 (such as *kas~ar*) and the Measure VII perfective active pattern $nC_1aC_2aC_3$ (such as *nkasar*) are derived from the perfective active pattern of the basic triliteral measure I $C_1aC_2aC_3$ (such as *kasar*), but at the same time geminating the second radical letter in C_1aC_2 ~aC_3 (*kas~ar*) and prefixing the derivational stem-prefix {*n*−} in $nC_1aC_2aC_3$ (*nkasar*). Notice here that the derived patterns share with the basic pattern the same final C_2VC_3 syllable (*sar*).

c) The *ta*-prefixed verbal measures (V, VI, and QII) also change the vowel of the final CVC syllable of the basic measure I/QI active voice imperfective pattern ($C_1C_2\mathbf{i}C_3$/ $C_1aC_2C_3\mathbf{i}C_4$) from '*i*' to '*a*' in the derived patterns. For example, we can get *takas~**ar*** (Imperf Act, Measure V) and *tadaHraj* (Imperf Act, Measure QII) but not **takas~**ir*** or **tadaHrij*.

The verbal inflectional paradigms also exhibit a number of syncretisms. To show these syncretisms, let us consider the perfective and imperfective paradigms. First, the following is the perfective paradigm formed from the root *ksr* using Measure II: First, Table 5.3 provides the perfective paradigm formed from the root *ksr* using Measure II. As the table shows, there are a number of syncretisms present in the suffixes:

- {*-tu*}, which indicates perfective first singular
- {*-naA*}, which indicates perfective first dual or plural
- {*-tumaA*}, which indicates perfective second dual

[4] The final syllable vowel (*i*) of the Measure I imperfective active pattern ($C_1C_2iC_3$) is changed to '*u*' or '*a*' according to stem root (see Table 5.1). For example, the imperfective active stem formed from the roots *ktb* and *ftH* using Measure I are ***ktub*** ($C_1C_2\mathbf{u}C_3$) and ***fta**H* ($C_1C_2\mathbf{a}C_3$) not * *ktib* and * *ftiH*.

Table 5.3. The perfective paradigm

	Mas	Fem
Perfective Active		
Singular		
1st	kas~ar-tu	kas~ar-tu
2nd	kas~ar-ta	kas~ar-ti
3rd	kas~ar-a	kas~ar-at
Dual		
1st	kas~ar-naA	kas~ar-naA
2nd	kas~ar-tumaA	kas~ar-tumaA
3rd	kas~ar-aA	kas~ar-ataA
Plural		
1st	kas~ar-naA	kas~ar-naA
2nd	kas~ar-tum	kas~ar-tun~a
3rd	kas~ar-uw	kas~ar-na
Perfective Passive		
Singular		
1st	kus~ir-tu	kus~ir-tu
2nd	kus~ir-ta	kus~ir-ti
3rd	kus~ir-a	kus~ir-at
Dual		
1st	kus~ir-naA	kus~ir-naA
2nd	kus~ir-tumaA	kus~ir-tumaA
3rd	kus~ir-aA	kus~ir-ataA
Plural		
1st	kus~ir-naA	kus~ir-naA
2nd	kus~ir-tum	kus~ir-tun~a
3rd	kus~ir-uw	kus~ir-na

There is also a partial syncretism expressed by the incorporated (connected) subject pronoun ‘*aA*,’ which indicates the dual number. In this context, the incorporated pronoun ‘*aA*’ is used as a complete suffix, as in {-***aA***}, and as parts of suffixes, as in {-*at****aA***} and {-*tum****aA***}. The reader will note that the morphosyntactic feature lists, which correspond to the three suffixes, share the dual number. Another partial syncretism is expressed by the incorporated subject pronoun ‘*at*,’ which indicates the third person singular feminine and which appears as a complete suffix ({-***at***}) and as part of a suffix ({-***at****aA*}).[5] A third partial syncretism in the perfective paradigm is expressed by ‘*t*,’ which indicates the second person in the suffixes {-***t****umaA*}, {-***t****a*}, {-***t****i*}, {-***t****un~a*}, and {-***t****um*}. All these suffixes share the second person.

[5] Notice that ‘*at*’ constitutes a third person dual feminine suffix when it is used with the dual ‘*aA*’ in the suffix {-***at****aA*}.

Table 5.4. The imperfective paradigm

	Mas	Fem
Imperfective Active		
Singular		
1st	Âa-ksir-u	Âa-ksir-u
2nd	ta-ksir-u	ta-ksir-iyn
3rd	ya-ksir-u	ta-ksir-u
Dual		
1st	na-ksir-u	na-ksir-u
2nd	ta-ksir-aAn	ta-ksir-aAn
3rd	ya-ksir-aAn	ta-ksir-aAn
Plural		
1st	na-ksir-u	na-ksir-u
2nd	ta-ksir-uwn	ta-ksir-na
3rd	ya-ksir-uwn	ya-ksir-na
Imperfective Passive		
Singular		
1st	Âu-ksar-u	Âu-ksar-u
2nd	tu-ksar-u	tu-ksar-iyn
3rd	yu-ksar-u	tu-ksar-u
Dual		
1st	nu-ksar-u	nu-ksar-u
2nd	tu-ksar-aAn	tu-ksar-aAn
3rd	yu-ksar-aAn	tu-ksar-aAn
Plural		
1st	nu-ksar-u	nu-ksar-u
2nd	tu-ksar-uwn	tu-ksar-na
3rd	yu-ksar-uwn	yu-ksar-na

Syncretisms in the imperfective paradigm can be seen by considering the imperfective paradigm formed from the root *ksr* using Measure I, listed in Table 5.4. As this table shows, there are a number of syncretisms present in the prefixes:

- { *ÂV*-}, which indicates imperfective first singular
- {*nV*-}, which indicates imperfective first person
- {*tV*-}, which generally indicates imperfective second person
- {*yV*-}, which indicates imperfective third person.

Additionally, there are syncretisms among the imperfective suffixes, which are evident in the following:

- {-*u*} , which corresponds to half the imperfective morphosyntactic feature lists
- {-*aAn*}, which indicates imperfective dual
- {-*uwn*}, which indicates imperfective masculine plural
- {-*na*}, which indicates imperfective feminine plural

Table 5.5. The imperative paradigm

	Mas	Fem
Singular		
2nd	kas~ir	kas~ir-iy
Dual		
2nd	kas~ir-aA	kas~ir-aA
Plural		
2nd	kas~ir-uw	kas~ir-na

The above paradigm also shows that the incorporated subject pronouns '*aA*' (dual) and '*uw*' (Mas Plural) have been used as part of the imperfective suffixes {-***aA**n*} and {-***uw**n*}. Hence, they represent a kind of partial syncretism here.

Finally, we turn to the imperative paradigm. Table 5.5 lists the imperative inflectional paradigm formed from the root *ksr* using Measure II. In this paradigm, we observe that there is a syncretism represented by the suffix {-*aA*}, which indicates the imperative dual. We also observe, when we compare this paradigm with the perfective paradigm, that there are syncretisms between the imperative and perfective paradigms represented by the suffixes (incorporated subject pronouns) {-*uw*} (Mas Plural), {-*na*} (Fem Plural), and {-*aA*} (Dual), which correspond to the following morphosyntactic feature lists:

- {-*uw*}: Imp 2nd Plural Mas, Perf 3rd Plural Mas
- {-*na*} : Imp 2nd Plural Fem, Perf 3rd Plural Fem
- {-*aA*} : Imp 2nd Dual Mas/Fem, Perf 3rd Dual Mas

When we compare the imperative paradigm with the imperfective paradigm, we see that the incorporated subject pronouns '*aA*'and '*uw*' represent a kind of partial syncretism if we consider the imperative suffixes {-***aA***}and {-***uw***} on the one hand and the imperfective suffixes {-***aA**n*} and {-***uw**n*} on the other. The incorporated subject pronouns '*aA*'and '*uw*' in the imperative and imperfective suffixes indicate Dual and Mas Plural, respectively. We also notice that the imperative suffix (incorporated subject pronoun) {-*na*} (Imp 2nd Plural Fem) represents a syncretism with the imperfective suffix {-*na*} (Imperf 3rd Plural Fem). The two suffixes indicate Plural Fem.

5.3 The Computational Approach

In a good computational approach to Arabic verbal morphology, the above generalizations, dependencies, and syncretisms should be employed. Capturing such generalizations, dependencies, and syncretisms in a computational implementation of Arabic morphology can save the system from redundancy and make it more compact.

To show how such generalizations, dependencies, and syncretisms can be used, in this chapter, I will introduce a computational approach to Arabic verbal morphology, which is constructed with DATR.[6] DATR is a lexical knowledge representation language based on the inheritance technique.[7] The approach will capture the above generalizations, dependencies, and syncretisms about Arabic verbal morphology in a compact non-redundant manner. The approach adopts the multilinear formalization of Arabic morphology introduced by McCarthy (1981, 1982), which organizes Arabic verbal stem structure using multiple layers of representation known as tiers. This approach has been also applied in Al-Najem (1998) to capture generalizations, dependencies, and syncretisms about Arabic nominal morphology using DATR.

The abovementioned generalizations will be handled in the implementation by encoding the generalizations once they are in generalization nodes higher in the inheritance hierarchy (inheritance network). Then, using default inheritance, other nodes lower in the inheritance hierarchy inherit, by default, information from these generalization nodes, and override, in some cases, some of the inherited default information. Thus, nodes lower in the inheritance hierarchy will inherit generalization information, which has been encoded once higher in the hierarchy, without re-encoding this information.

The dependencies between verbal measures will be systematized through stem partitioning and through the multiple inheritance technique. In this context, a node in this implementation that represents a derived (dependant) measure will inherit a shared stem portion (the final CVC syllable) from the node representing its corresponding original measure. The node representing the derived form defines the remaining portion, which precedes the final CVC shared portion.

Syncretisms in inflectional paradigms are handled mainly using default inference technique. Using a default interference technique, the implementation can infer, from a single (usually underspecified) statement about the inflectional paradigm, more statements about that paradigm. These statements are implicitly inferred, by default, from the original single statement without re-encoding them explicitly.

The following is a demonstration of how generalizations are handled in the implementation. The abovementioned generalizations about verb inflection will be encoded in our implementation in the node `Verb`, which contains the following information

```
Verb:

<$vform syn cat> == verb

<$vform mor stem> == "<$vform p l>" "<$vform p r>"

<$vform mor stem imp> == "<$vform mor stem imperf act>"

<$vform mor> == "<$vform mor stem>" Verb_Infl_Aff:<mor
```

[6] For more details about the computational approach introduced in this chapter and for details on other computational approaches to Arabic morphology, see Al-Najem (1998).

[7] For an introduction to DATR, see Evans (1990) and Evans and Gazdar (1990, 1996).

```
suf>
<$vform mor imperf> == Verb_Infl_Aff:<mor pre imperf>
Imperf_Pre_V:<a> "<$vform mor stem imperf>"
Verb_Infl_Aff:<mor suf imperf>
<$vform mor imperf pass> == Verb_Infl_Aff:<mor pre
imperf pass> Imperf_Pre_V:<u> "<$vform mor stem imperf
pass>" Verb_Infl_Aff:<mor suf imperf pass>.
```

Notice here that the imperfective prefix, in this implementation, consists of an imperfective prefix consonant inherited from the node `Verb_Infl_Aff` and an imperfective prefix vowel inherited from the node `Imperf_Pre_V`. This allows for the generalization that the imperfective active prefix vowel is normally '*a*' but is changed to '*u*' when imperfective active prefixes are used with imperfective active stems of Measures II, III, IV, and QI. In addition, as previously indicated, the passive imperfective prefixes are the same as the active imperfective prefixes except that the vowel '*u*' is used instead of '*a*.' The imperfective prefix begins with a consonant (*'n', 'Â', 'y'*, or '*t*'), which has no variation with respect to a measure or a voice (imperfective active/passive). The variation in the imperfective prefixes occurs just to the prefixes' vowel ('*a*' or '*u*'), as we have already seen, according to the measure and the voice. Thus, instead of explicitly encoding the prefixes with the two variations of vowel, which yields two sets of prefixes: {*na, Âa, ya, ta*} and {*nu, Âu, yu, tu*}, we have only encoded the consonants of the prefixes in the node `Verb_Infl_Aff`, and the vowel is chosen by inheritance from the node `Imperf_Pre_V`.

The sixth statement of `Verb` partially overrides the third statement, declaring that the imperfective passive inflection paradigms are constructed by prefixing an imperfective passive prefix ending with the vowel '*u*' to an imperfective passive stem and suffixing an imperfective suffix. The fifth and sixth statements, using default interference, provide the imperfective active and passive paradigms. In other words, they allow the generation of all the 36 active and passive imperfective verb forms.

The node `Verb`, which encoded generalizations about verbal paradigm formation, occurs high in the inheritance hierarchy. It will be inherited, by default, several times by other nodes lower in the inheritance hierarchy without re-encoding the generalizations of `Verb` again in those nodes. Deault inheritance permits us to encode these linguistic generalizations in a compact, non-redundant manner. Note also that the use of default inference in the last three statements of `Verb` is crucial to allow the generation of 78 verb forms, that is, virtually all the Arabic morphology verb forms.

The lower nodes inheriting from `Verb` represent measures, and these nodes produce the inflectional stems[8] that correspond to each measure from radical letters inherited globally from root nodes. These stems will be used in the formation of the inflectional paradigms. A difficulty could arise here. The node `Verb` states that the (default) way of constructing the imperfective active paradigm is by prefixing an imperfective active prefix ending with the vowel '*a*' to an imperfective active stem and suffixing an imperfective suffix. This holds for most of the measures. However, some measures construct the imperfective active paradigm by prefixing an imperfective active prefix ending with the vowel '*u*' instead of '*a*.' These are Measures II, III, IV, and QI, as previously stated. The use of the vowel '*u*' in imperfective active prefixes represents a sub-generalization (sub-regularity) emerging from the generalizations (regularities) of the node `Verb`, which encoded the generalizations previously stated in (1). So, we need a way to define this sub-generalization in a node that can be inherited by specific nodes representing Measures II, III, IV, and QI. This is achieved by defining an intermediate node[9] (`Verb2`), which inherits, by default, all the information of the default (regular) node `Verb` but overrides the inherited statement that encodes the imperfective active paradigm by changing the original prefix vowel '*a*' to '*u*.'

```
Verb2:

<> == Verb

<$vform mor imperf> == Verb_Infl_Aff:<mor pre imperf>

Imperf_Pre_V:<u> "<$vform mor stem imperf>"

Verb_Infl_Aff:<mor suf imperf>.
```

Thus, while the node `Verb` represents the default way of constructing the verbal paradigms, `Verb2` represents the sub-generalization (sub-regularity) that is related to the vowel of the imperfective active prefixes used with the imperfective active stems of Measures II, III, IV, and QI.

So far, we have considered how generalizations about verb formation can be encoded in our implementation. Now, we turn to generalizations relating to the dependencies between verbal measures. Recall in this context, as stated in (2), that the patterns of most of the triliteral verbal measures are normally derived from the corresponding patterns of the triliteral verbal measure I by means of changing the pattern portion that precedes the final C_2VC_3 syllable of these derived patterns. In addition, most of the patterns of the verbal quadriliteral measures are derived from the corresponding patterns of the verbal quadriliteral measure QI by changing

[8] Examples of these inflectional stems are the imperfective active and imperfective passive stems.

[9] The use of intermediate nodes is a common way adopted in DATR literature to capture sub-regularities. See Corbett and Fraser (1993) for example.

the pattern portion that precedes the final C_3VC_4 syllable of these derived patterns. This means that to capture the dependencies between the verbal measures, we need to split the verbal stem (pattern) into two parts: the final CVC syllable, which is normally invariably shared between the stems (patterns) of the derived and the original measures and a remainder, which is the part of the stem (pattern) preceding the final CVC syllable.[10] To achieve this, the verbal stem in our implementation will consist of two parts: a right part that represents the final CVC syllable of the stem and a left part that represents the remainder of the stem preceding the final CVC syllable.

To demonstrate how the dependencies between verbal measures are captured in our implementation, we take the following example, involving the dependencies between Measures III (*faAçal*) and VIII (*ftaçal*) on the one hand and the basic verbal Measure I (*façal*) on the other hand. The following is a definition of the node `Façal`, which represents the basic verbal Measure I:

```
% The CV-patterns of the Measure I (façal)

Façal:

<> == Verb

<façal p l> == C:<1> V:<>

<façal p l imperf> == C:<1>

<façal p r> == C:<2> V:<*> C:<3>.
```

The node V below encodes the consonants of a root through global inheritance from nodes representing roots, and the node C encodes perfective and imperfective vocalisms:

```
%The Vocalism Tier
V:

<perf act> == Perf_Act:<>

<perf pass> == Perf_Pass:<>

<* perf act> == <perf act *>

<* perf pass> == <perf pass *>
```

10 This remainder is not always a single complete free standing syllable. It may become part of a syllabic structure through the insertion of the epenthetic '*ÂV*,' as in **(Âi)n**$\mathbf{C_1}$**a** C_2aC_3 (Measure VII, Perf Act), which yields the syllabic structure **CVCCV**. It may also become part of a syllabic structure by the prefixation of an inflectional prefix like {*ya-*} as in **(ya)**$\mathbf{C_1}$ C_2iC_3 (Measure I, Imperf Act), which yields the syllabic structure CVC.

```
<imperf act> == Imperf_Act:<>

<imperf pass> == Imperf_Pass:<>

<* imperf act> == <imperf act *>

% "<* ... >" represents a terminal non-default vowel


%Vocalism Generalizations

Perf_Act:

<> == a.

Perf_Pass:

<> == u

<*> == i.


Imperf_Act:

<> == a

<*> == i.



Imperf_Pass:

<> == a.



% The association of the root tier elements (radical

% letters) to the C slots of the pattern tier

# vars $n: 1 2 3 4.

C:

<$n> == "<$n>".
```

The node `Façal` encodes the right parts (final CVC syllables) and the left parts, which correspond to the perfective and imperfective stems (patterns) of Measure I. Notice also that we have not encoded the right and left parts of the imperative stems since the imperative and imperfective active stems are the same as is indicated by the statement (generalization):

```
<$vform mor stem imp> == "<$vform mor stem imperf act>"
```

of the node `Verb`, which is inherited, by default, by the current node. This generalization about imperative stems has been encoded once in `Verb` higher in the inheritance hierarchy, and it is then inherited by the measure nodes lower in the hierarchy. The right and left parts of the perfective and imperfective stems are combined together to form complete perfective and imperfective stems by one statement encoded once in the node `Verb` higher in the inheritance hierarchy, which is inherited, by default, by Façal and the measure nodes lower in the hierarchy:

```
<$vform mor stem> == "<$vform p l>" "<$vform p r>"
```

This line states that a verbal stem consists of a left part followed by a right part. Default inference will fill the gaps, allowing the formation of all the perfective and imperfective stems. For instance, this single, short, and general line will (implicitly) stand for longer more specific lines like the following, which constructs a perfective active stem:

```
<$vform mor stem perf act> == "<$vform p l perf act>"

"<$vform p r perf act>"
```

Then, the node `FaAçal` inherits the right parts (final CVC syllables) of `Façal`, but it defines its left parts in a different way:

```
% The CV -- patterns of the Measure III (faAçal)

FaAçal:

<> == Verb2

<faAçal p r> == Façal:<façal p r>

<faAçal p l> == C:<1> V:<> V:<>.
```

The underspecified line:

```
<faAçal p l> == C:<1> V:<> V:<>
```

states, using default inference, that the (default) left part of all the perfective and imperfective stems (patterns) of `FaAçal` is constructed as the first radical letter of a root followed by a long vowel. The vowel is inherited from the vocalism node `V`.

In addition, the node `FaAçal` inherits, by default, the generalizations encoded in `Verb2`.[11] The reader will, therefore, notice that in this node, we adopt a multiple inheritance by allowing the node `FaAçal` to inherit from two nodes, `Verb2` and `Façal`.

Thus, in the node `FaAçal`, we have captured the dependency between Measures III (*faAçal*) and I (*façal*) using multiple inheritance. We inherited the generalizations about verbs from `Verb2` and inherited the invariant right parts (final CVC syllables) that are shared with Measure I from the node `Façal`. Measure III (*faAçal*) is derived from I (*façal*) except that the left parts of the stems of *façal* are changed to C_1aA (*faA*) (and C_1uw in the perfective passive). In other words, the dependency generally involves the lengthening of the first vowel. Notice here that since we have encoded verb generalizations once in the higher node `Verb`, from which the intermediate node `Verb2` inherits, we do not need to re-encode these generalizations again in the node `FaAçal`. We have simply inherited these generalizations, by default, from `Verb2`. Notice also that we have not re-encoded the right parts (the final CVC syllables) of the stems of `FaAçal` since we encoded these parts once in `Façal` and passed them on from this node. Defining the left parts of `FaAçal` is accomplished in a compact non-redundant manner by using default inference through underspecifying the left part using the statement:

```
<faAçal p l> == C:<1> V:<> V:<>
```

This means that this statement, using default inference, will represent all the left parts of all the perfective and imperfective stems of `FaAçal`. Thus, the dependency between measures III (*faAçal*) and I (*façal*) has been captured by means of multiple inheritance in a compact non-redundant manner, without the need to repeat encoding information which has previously been encoded in higher nodes.

In a similar fashion, the dependency between Measure VIII (*ftaçal*) and I (*façal*) can be captured:

```
% The CV-patterns of the Measure VIII (ftaçal)

Ftaçal:

<> == Verb

<ftaçal p r> == Façal:<façal p r>

<ftaçal p l> == C:<1> Affix:<t> V:<>.
```

This node illustrates that the stems of Measure VIII (*ftaçal*) are derived from the stems of Measure I (*façal*) by inserting the infix {*t*} after the first radical letter in the

[11] As previously seen, the node `Verb2` is a sub-generalization intermediate node, which inherits, by default, the information of the generalization node `Verb`.

left part of the stems of *ftaςal*. The node `Affix` defines derivational affixes used in the CV-pattern tier and is defined below:

```
Affix:

<t> == t

<n> == n

<st> == s t

<Â> == Â

<w> == w.
```

The affixation of derivational affixes is indicated in the pattern tier by inheritance descriptors like `Affix:<t>`, which indicates (inherits) the reflexive infix $\{t\}$.

The other dependencies between verbal measures are captured using a similar method to that used with the dependencies between the previous verbal measures. The inheritance hierarchy below demonstrates how the dependencies between verbal measures are captured in our implementation:

Note that some derived measures inherit from multiple nodes (multiple inheritance). Multiple inheritance is indicated in this diagram by having a derived measure node which inherits from more than one node (receiving more than one IN arrow).

So far, we have considered how generalizations relating to verb formation and generalizations about the dependencies between verbal measures are captured in our implementation. We will now discuss syncretisms in inflectional paradigms and how these are encoded in our DATR implementation.

The verbal inflectional paradigms are handled in our implementation by means of the node `Verb_Infl_Aff`. In this node, we define paths representing morphosyntactic feature lists and assign these paths values representing inflectional affixes (prefixes and suffixes) that correspond to the paths' morphosyntactic feature lists, as we will see shortly. The inflectional paradigms themselves are built through the generalization nodes `Verb` and `Verb2` using statements like the following statements taken from the definition of the node `Verb`:

```
<$vform mor> == "<$vform mor stem>" Verb_Infl_Aff:<mor

suf>

<$vform mor imperf> == Verb_Infl_Aff:<mor pre imperf>

Imperf_Pre_V:<a> "<$vform mor stem imperf>"

Verb_Infl_Aff:<mor suf imperf>

<$vform mor imperf pass> == Verb_Infl_Aff:<mor pre
```

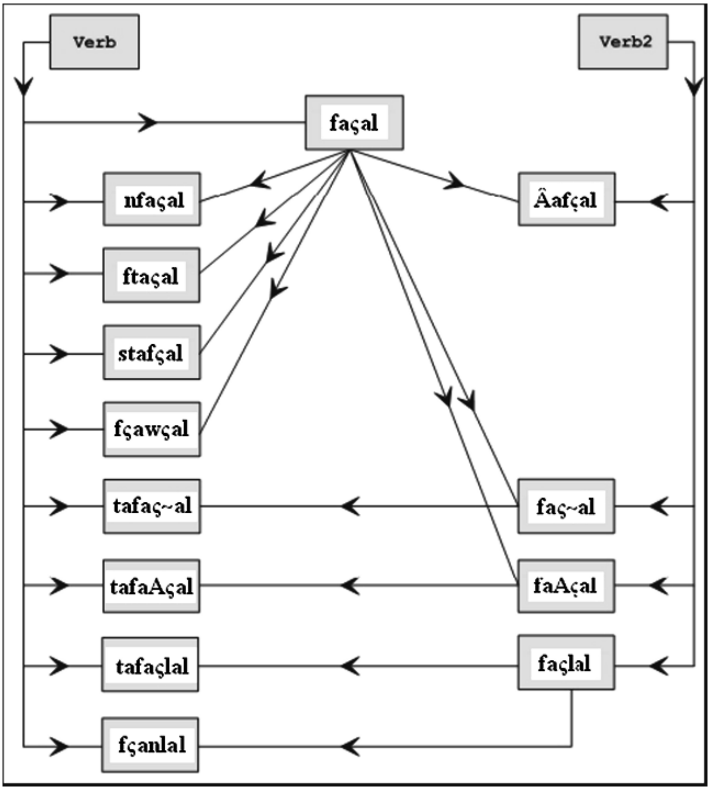

Fig. 5.1. An inheritance hierarchy showing dependencies between verbal measures

```
imperf pass> Imperf_Pre_V:<u> "<$vform mor stem imperf

pass>" Verb_Infl_Aff:<mor suf imperf pass>
```

These three statements, by inheriting the affixes from `Verb_Infl_Aff` and especially through default inference, allow the construction of the whole verbal paradigm of Arabic morphology in a compact, non-redundant manner, as we have already seen from consideration of the node `Verb`.

In our implementation, syncretisms in paradigms are handled mainly using default inference. Default inference represents a useful means of expressing generalizations and encoding repeated information in a compact non-redundant manner. It saves us from re-encoding information which can be inferred, by default, from one (usually underspecified) path. Consider in this context the statement:

```
<mor suf imperf act> == u
```

of the node `Verb_Infl_Aff`, which is the node where inflectional verbal affixes of the inflectional paradigms were encoded. From this single statement, which is underspecified for person, number, and gender, we can infer that, roughly, the default suffix of the imperfective active paradigm is {-*u*}. This is because the suffix {-*u*} is associated with half of the morphosyntactic feature lists of the imperfective active paradigm. Then, we partially override this statement to encode the other imperfective active suffixes, which are associated with the remaining morphosyntactic feature lists of the imperfective active paradigm. From the previous single path (`<mor suf imperf act>`), we can infer, by default, other more specific paths, which include the following:

```
<mor suf imperf act first sing mas>

<mor suf imperf act first dual mas>

<mor suf imperf act first plural mas>
```

These paths are implicitly inferred, by default, from the previous single underspecified path without re-encoding them explicitly. This is a better method than others like local inheritance from (referral to) a specific local path, as in the following:

```
<mor suf imperf act first sing mas> == u

<mor suf imperf act first dual mas> == <mor suf
imperf act

first sing mas>

<mor suf imperf act first plural mas> == <mor suf
imperf act

first sing mas>
```

which is not a compact way in comparison to default inference.

Thus, the previous example clearly shows how default inference represents a useful means to capture syncretisms in paradigms in a compact non-redundant manner. Using a single short (usually underspecified) path, we can infer, by default, multiple longer more specific paths without re-encoding these inferred paths and their values explicitly.

As previously noted, there are some partial syncretisms in the verbal inflectional paradigms, which can be captured in the implementation. Partial syncretisms mainly concern incorporated subject pronouns, which materialize as part of suffixes or as complete suffixes. To handle partial syncretisms in our implementation, we encoded (generalizations about) such incorporated pronouns once in a node called `Part_Sync`:

```
Part_Sync:

<dual> == a A
```

```
<plural mas> == u w

<plural fem> == n a

<sing fem> == a t

<second> == t.
```

Using this node, such incorporated pronouns will not be re-encoded explicitly in the node `Verb_Infl_Aff`, which encodes the syncretism in verbal inflectional paradigms. Instead, we refer to the node `Part_Sync` to inherit those incorporated pronouns, which represent partial syncretisms. An example that shows this is the following statement from the node `Verb_Infl_Aff`:

```
<mor suf imperf act $2nd_3rd $dual_plural> ==

Part_Sync:<$dual_plural> n
```
[12]

Through this statement, the incorporated pronouns '*aA*' (Dual) and '*uw*' (Plural Mas) are inherited from `Part_Sync`. The incorporated pronouns '*aA*' and '*uw*' are parts of the suffixes {***-aA****n*} and {***-uw****n*} associated with the morphosyntactic feature lists :

- {*-aAn*} :

Imperf Act 2nd Dual Mas, Imperf Act 3rd Dual Mas, Imperf Act 2nd Dual Fem, and Imperf Act 3rd Dual Fem.

- {*-uwn*} :

Imperf Act 2nd Plural Mas, and Imperf Act 3rd Plural Mas.
The previous statement encodes this using the node `Part_Sync` and default inference. This statement is partially overridden to encode the other suffixes which are associated with the imperfective active second singular feminine in addition to the imperfective active second and third plural feminine. At the same time, the incorporated pronouns '*aA*' and '*uw*' (and other incorporated pronouns) also form complete suffixes, as can be seen in the following statement from the node

```
Verb_Infl_Aff:

<mor suf perf act third> == Part_Sync:<>
```

[12] The variables `$2nd_3rd` and `$dual_plural` have the following definitions, respectively:

```
# vars $2nd_3rd: second third.
# vars $dual_plural: dual plural.
```

Using this single statement, the incorporated pronouns '*aA*' and '*uw*', in addition to other incorporated pronouns, will form complete perfective active third person suffixes corresponding to morphosyntactic feature lists, including Perf Act 3rd Dual Mas and Perf Act 3rd Plural Mas. The single underspecified paths of that statement will stand for paths such as:

```
<mor suf perf act third dual mas> == Part_Sync:
<dual mas>

<mor suf perf act third plural mas> == Part_Sync:
<plural mas>
```

in addition to others using default inference.

In addition to the above nodes, which define how verbal stems and inflectional paradigms are constructed, the implementation contains other nodes known as the lexical nodes. The lexical nodes are those representing roots and defining the radical letters of these roots. The radical letters will be inherited, using global inheritance, by other nodes in the implementation representing verbal measures to form verbs. The lexical nodes also inherit (select) verbal measures, which can be used to form verbs using the roots represented by those lexical nodes. These nodes are the query nodes of our implementation, and they represent the initial (global) context of inheritance. The queries that will be entered into the implementation will be queries about these lexical nodes. An example of the lexical nodes is the node `KSR`, which represents the root *ksr*.

```
KSR:

<c1> == k

<c2> == s

<c3> == r

% Verb forms

<façal> == Façal

<faç~al> == Faç~al

<nfaçal> == Nfaçal

<tafaç~al> == Tafaç~al
```

So, using a node like this one and through the use of the other abovementioned nodes, we can enter queries like the following and obtain their corresponding results:

```
| ? datr_theorem ('KSR', [façal, mor, imperf, act,
third, dual, mas]).
yaksiraAn
```

```
| ? datr_theorem ('KSR' [faç~al, mor, perf, act, first,
sing, mas]).
kas~artu
```

```
| ? datr_theorem ('KSR' [tafaç~al, mor, perf, act,
third, sing, mas]).
takas~ara
```

```
| ? datr_theorem ('KSR' [façal, mor, perf, pass, third,
sing, mas]).
kusira
```

```
| ? datr_theorem ('KSR', [faç~al, mor, imp, second,
 sing, fem]).
kas~iriy
```

DATR was originally used for generation (synthesis) only, not for recognition (analysis). Information is obtained from DATR theories via queries that cause DATR to *generate* results which take forms like theorem dumps.

5.4 Conclusion

In this chapter, we have computationally systematized generalizations, dependencies, and syncretisms existing in Arabic verbal morphology in a compact, non-redundant manner. Generalizations have been handled through the use of DATR default inheritance, as seen in the node `Verb`. Dependencies have been handled through the use of DATR multiple inheritance, as seen in nodes like `Façal` and `FaAçal`. Finally, syncretisms have been handled through the use of DATR default inference, as seen in the node `Verbs_Infl_Aff`. As has been mentioned earlier, capturing such generalizations, dependencies, and syncretisms in a computational implementation of Arabic morphology can save that implementation from redundancy and can make it more compact.

References

Al-Najem, Salah. "*Computational Approaches to Arabic Morphology*." Diss. Essex University, 1998.

Corbett, Greville and Norman Fraser. "Network Morphology: a DATR A Account of Russian Nominal Inflection." *Journal of Linguistics* 29 (1993): 113–142

Evans, Roger. "An Introduction to the Sussex PROLOG DATR System." *The DATR Papers*. Eds. Roger Evans and Gerald Gazdar . Brighton: University of Sussex, 1990. 63–70.

Evans, Roger and Gerald Gazdar, eds. *The DATR Papers*. Vol. 1. Brighton: University of Sussex, 1990.

Evans, Roger and Gerald Gazdar. "DATR: A Language for Lexical Knowledge Representation." *Computational Linguistics* 22.2 (1996): 167–216.

McCarthy, John. "A Prosodic Theory of Nonconcatenative Morphology." *Linguistic Inquiry* 12 (1981): 373–418.

McCarthy, John. "*Formal Problems in Semitic Phonology and Morphology*." Diss. Indiana University, 1982.

Trask, R. L. *A Dictionary of Grammatical Terms in Linguistics*. London: Routledge, 1993.