

Aspects of Phonology / Morphology Interface in Arabic

DENDANE Zoubir
Université de Tlemcen

Résumé:

Cette communication tente d'explorer quelques aspects de l'interaction entre deux niveaux linguistiques dans la langue arabe, la phonologie et la morphologie. Cette interface qui révèle des caractéristiques très subtiles et hautement complexes dans les structures grammaticales de cette langue, a suscité des recherches sophistiquées qui ont fait avancer les théories. L'intérêt porté sur l'arabe par les linguistes et les informaticiens résulte du fait que cette langue possède un système morphologique ayant des traits inexistant dans d'autres langues, ce qui présente un défi partagé par la recherche en morphologie universelle, d'un côté, et en linguistique computationnelle, de l'autre.

1. Introduction

The present paper attempts to explore a few aspects of the interaction between two linguistic components or levels of the Arabic linguistic system, phonology and morphology. Such interface has not only revealed some subtle and highly complex characteristics in the grammar of this language, but also driven many researchers toward far-reaching investigations and theoretical implications. These new developments often bring explanations to

long-lasting problems, but also expose new issues at a most abstract and tantalizing level. Indeed, the complex morphological structure of Arabic, based primarily on non-concatenative inflectional and derivational patterns, in addition to concatenative affixation, has certainly had a powerful incentive for the development of recent morphological theories, including, in particular, root-and-pattern morphology (McCarthy 1979), but also prosodic morphology (McCarthy and Prince 1986 *et seq.*) and two-level morphology (Koskeniemi's 1983)¹. One reason for the interest shown by linguists and computer scientists for Arabic is that this natural language is characterized by morphological features non-extant in other languages.

Our goal in this communication is first to provide a brief picture of the variety of shapes taken by Arabic words, particularly in verb moulds, or *âwzaan*, whose organisation is based on a templatic system that presents motivating challenges to universal morphology, on the one hand, and to computational linguistics, on the other.

Then, section two introduces a few patterns which show how underlying phonological representations and morphological processes are mutually influenced. The talk will also touch upon some principles proposed in Prosodic Morphology, described in McCarthy and Prince (2003:1) as "...a theory of how morphological and phonological determinants of linguistic form interact with one another in a grammatical system."

In the last section, we will look at some instances of phonological/morphological interface in Arabic in the

¹ Actually, Koskeniemi's *two-level morphology* encompasses what is considered by linguistics as morphology but also phonology, in the sense of morphophonemics, at least.

light of the *Optimality Theory* framework proposed in Prince and Smolensky (1993, *et seq.*).

2. Aspects of Arabic Morphology

The Arabic language is known for its highly inflexional morphology, particularly illustrated by vowel infixation, a process wherein vowel quality is used to distinguish between morphological categories. In this process, basic consonantal roots² and vowel ‘melodies’ are combined according to a predetermined number of canonical shapes to generate fixed patterns, also called ‘templates’ (Cf. McCarthy’s 1979), used to express the various lexico-morphological categories.

2.1 Arabic Verbs

The Arabic verb, which has played a predominant role in the language, is formed on the basis of a tri-consonantal root-morpheme – referred to by traditional Arabic grammarians as *maadda*, meaning ‘matter’ or ‘substance’ – interspersed with what may be considered a vowel morpheme or a ‘melody’ consisting of vocalic elements called *ʿarakaat*³ (literally, ‘movements’, but meaning here vowels that make consonants move). The two morpheme sets are arranged, or rather intercalated,

² These are not to be viewed as roots in the traditional morphological sense, as they only consist of three consonants (mainly) interspersed with a vowel ‘melody’ that produces a given template or category

³ The short vowels *a*, *u*, *i* are not usually represented in written text, except in the Koran or for other expository purposes such as text books for elementary school pupils.

according to fixed patterns and word configurations that make Arabic morphology of the non-concatenative type. Such a complex and regular process allows for the generation of a large number of basic verbal variants from which derived forms are obtained by means of affixation. In an attempt to set up a programme based on a computational treatment of Arabic morphology, Soudi *et al.* (2001:3) describe the processes of affixation as follows:

In addition to prefixation and suffixation, inflectional and derivational processes may cause stems to undergo infixational modification in the presence of different syntactic features as well as certain stem consonants.

Following the conventional root {f.ʕ.l}, the morpheme {ktb}, which carries the notion of writing, produces various patterns according to the vowel melody applied:

- In the perfective, {a-a} produces the template /katab/⁴ in the active form ‘he wrote’, while passivization obtains by way of the regular vowel pattern {u-i} giving /kutib/. Reciprocity, on the other hand, is expressed by means of the vocalic sequence {aa-a} that yields /kaatab/, ‘to write to each other’. It is worth noting that in perfect stems, three types of basic form I verbs obtain according to the vowel tier: {a-a} is found in /daʕal/ or

⁴ The full form of the perfective verb, which is equivalent to the 3rd pers. sing. paradigm, is said to be ‘built on the open (vowel)’ in Arabic (*mabniy ʿalal fatḥ*); that is, it ends with the vowel /a/, giving /kataba/. It is thus not considered here for the redundancy.

/ʔakal/, ‘enter’ and ‘eat’, {a-i} in verbs like /laʕib/, ‘play’ and {a-u} in forms like /kabur/, ‘It grew larger’.

- In the imperfective, whose base template takes the form CVCCVC, /jaʕʕal/, the active/passive distinction is achieved by means of the inversion of the vocalic pattern, i.e. /jaktub/ vs. /juktab/, ‘He writes vs. It is written’.

The process of infixation is not involved in the construction of imperfective stems; these have the regular pattern /ktub/, the paradigms of which are rendered by pronoun prefixation as in /ʔaktub/, /taktub/, /jaktub/, /naktub/ for ‘I write, You write, He writes, We write’, respectively, and suffixes that represent the dual, plural forms (masculine and feminine) like /jaktubaan(i)/, /jaktubuun(a)/ and /jaktubn(a)/, respectively.

The overall Arabic verbal morphology thus shows, in subtle ways, the relationship between verb templates and syntactic operations, on the one hand, as in the case of passivization in which melodic overwriting is enough to convey forms like /ʔukil/ and /juʔkal/ ‘It has been/was eaten’ vs. ‘It is eaten’. Lexico-phonological interfering, on the other hand, can also be shown in cases of reciprocalization where phonological elements are added in the form of infixes as in /kaatab/, ‘to correspond (by writing)’, or prefixes as in /ʔistaktab/ ‘to ask someone to write’, to obtain diverse but semantically related lexical items. But in the morphological shape of form II /faʕʕal/, as in verbs like /qattal/, meaning something like ‘to overkill or to exaggerate in killing’⁵, the C₂ gemination,

⁵ The base form I /qatal/, without *ʕaddah*, means ‘to kill’.

represented by the *ḡaddah* – in Arabic, clearly embodies intensity which conveys a semantic feature of the form. Other types of verbs which display attractive cases of morpho-phonological interaction are those consisting of ‘weak’ radicals which we touch upon in the next sections along with other interesting instances of phonological/morpho-syntactic interface.

2.2 Arabic Nouns

Arabic nouns too – as well as adjectives – are governed by a complex set of canonical patterns which reveal phonological/morphological interactions that are worthy of attention, in particular because of the fact that many are usually derived from verbal consonantal roots. In addition, the plural of many nouns, especially in the masculine, is not formed by suffixation, as it occurs with what is referred to as ‘sound plurals’ to which the suffix morphemes masc. {-uun} or fem. {-aat} are attached, e.g: ‘teachers’ → masc. /muʕallimuun/; fem. /muʕallimaat/. ‘Broken plurals’, much more widespread in the language, have a more complex and more appealing morphological structure because of the internal modification they are characterized with. The modification of the singular stems occurs in a great number of canonical plural patterns according to some words’ prosodic traits. In words like /kitaab/, the stem itself is subjected to change by means of vowel melody alteration yielding the plural form /kutub/, ‘books’.

Also, the very common noun pattern CVCC, for instance, may have unpredictable plural forms: /kalb/ pluralizes as /kilaab/, ‘dogs’, with a CiCaaC pattern while /nažm/, ‘star’, with the same CVCC pattern, has a

CuCuuC form, /nuʒuum/. Furthermore, many nouns may have two to five plural allomorphs, e.g. /baħr/ → pl. /biħaar/, /ʔabħur/, /buħuur/, all three meaning ‘seas’. Some nouns may even have both sound and broken plurals; for example, ‘a capital city’ /ʕaaʕima/ pluralises as /ʕaaʕimaat/ and /ʕawaaʕim/. And to make things more complex, Arabic uses the ‘plural-of-the-plural’ pattern for many nouns like /ʒamal/, ‘camel’, whose plural is /ʒimaal/ with two plurals-of-the-plural, /ʒamaaʔil/ and /ʒimaalaat/.

A comprehensive categorization of pluralisation in Arabic is indeed so arduous, but so appealing at the same time, that it requires the devising of a computer-based bank of plural patterns, if it does not exist already. In fact, only with the development of a computational model that can handle Arabic morphology generation as a whole, on the basis of the recent theoretical proposals, can we reach a better appreciation of the Arabic grammar system in terms of the interference of its components.

We consider in the next sections cases that reveal how Arabic verb templates as well as noun patterns have to ‘behave’ in the overall grammar of the language. But it is essential and attention-grabbing at the same time to point up the important roles that short vowels play in the language. In fact, this is another specificity of Arabic that is illustrated in the functions assigned to vowels, as illustrated below.

2.3 Arabic Vowels

While consonants in Arabic appear to be associated with lexico-semantic concepts – {ktb} is the consonantal root

or substance for ‘writing’, {klm} for ‘speaking’ and so forth –, vowels are responsible for functions such as verb or noun categorization and syntactic marking, as we will attempt to show in the following.

- First, it is worth mentioning, *en passant*, the fact that the final short vowel of each word in a sentence and the ‘nunation’ (*at tanween*), a marker of indefiniteness, tend to be dropped in pause form, usually at the end of an utterance, even in the language of the Koran with its status as the most formal type of Arabic⁶. For example, in the Koranic verse /qul huwa llaahu ʔahadun/, the last word is read out [ʔahad], without *tanween*; and /ʔuktub ismaka/ is realised [ʔuktub ismak].

But this does not mean at all that final vowels have no function in the language. As a matter of fact, they are phonological elements serving as syntactic markers, for they are used to distinguish, for instance, between a subject that has to end with a *ʔamma* (a rounded vowel) and an object which must carry a final *fatʔa* (an open V); e.g. /ʔakala l waladu ttuffaaħata/, ‘The boy has eaten the apple’. Native speakers are of course supposed to have learned these vocalic patterns, or *iʔraab* rules, at school.

- Second, the first vowel V₁ in (non-augmented) verb forms in the indicative (perfect and imperfect) is always a *fatħa* in the active form and a *ʔamma* in the passive; this is a clear indication that the vowels /a/ and

⁶ Final vowel deletion is attested even within the utterance in MSA, and practically at the end of each word in dialectal Arabic.

/u/ in V₁ position are assigned the syntactic function of differentiating between the two forms, as in /ʔaakal/ vs. /ʔuukil/ (He ate vs. It was eaten) and /jaʔku/ vs. /juuʔku/. The first vocalic element of verbs, V₁, has indeed been referred to by Teeple (2005:3), and others before him, as “the active voice vowel”.

- Third, V₂, the second vocalic slot of the verb stem, is allocated one of the three vowels in accordance with overall semantic considerations; verbs of action, for instance, usually have a /faʕa/ form, like /daʕal/ ‘enter’ and /ʕaraʕ/ ‘go out’, in opposition with the /faʕu/ pattern often associated with semantic values related to state, process or quality. On the other hand, the template /faʕil/, “un verbe moyen”, as Larcher⁷ sees it, represents a form that conveys a sense somehow between action and state; e.g. /ʕarib/ ‘drink’ vs. /ʕalim/ ‘know’.

3. Phonology/Morphology Interface

Although individual components of a grammar can be – and are indeed – examined in isolation for specific purposes, it is now admitted that more insightful accounts can be obtained through the analysis of how such components are inter-related and how they function in harmony in language systems. The rules operating within a given level, as documented in traditional morphology and phonology (Chomsky and Halle 1968), may influence or be influenced by rules related to the ‘contiguity’ of grammar components.

⁷ In Dichy, 2003:344.

The point is that a fruitful exploration of the Arabic language requires that its grammatical organisation be considered as an entity, not in terms of separate levels. Bird and Blackburn (1994: 1) write in this respect:

One model which looks particularly attractive in this regard considers the traditional modules of grammar (*i.e.* syntax, semantics and phonology) not in *series* where the output of one feeds into the input of the next, but rather in *parallel*, where each module exerts independent constraints.

Linguistic structures are indeed better handled in terms of integrated entities to be analysed simultaneously from different perspectives, taking into account two or more levels at the same time.

A good example of such interface is obvious in the English plural morpheme {-s}: the allomorphic variation, *i.e.* [-s], [-z] and [-iz] as in *book[s]*, *boy[z]* and *bus[iz]*, are clearly accounted for with reference to phonological processes, here the preceding phonetic segments being the influencing elements; in *boys*, /s/ → [z] / [+voice] – . Such cases of assimilation may be regarded as instances that result from the natural and unconscious tendency of speakers to ‘make’ speech production as harmonious as possible.

Other relevant examples that reveal interactions between phonology and morphology are relatively easy to come across in languages. In fact, native speakers of any language are linguistically competent if they ‘know’ how to combine phonological, morphological and syntactic processes.

3.1 Vowel harmony

A similar allomorphy to the English plural {-s} occurs with the suffix pronoun {-hu} (possessive, objective case, etc.) in Arabic. In [kitaabu**hu**] ‘His book’, [daraba**hu**] ‘He hit him’ or [qul la**hu**] ‘Tell him’, {-hu} is realised as a {-hi} allomorph as a result of the high front vowel [i] or the unvowelled glide [j] that precedes [h], as in [bi**hi**], ‘with it’, and [ʔila**jhi**], ‘to him’, and not *[bihu] or *[ʔilajhu]. Here, the phenomenon is more complex than that of the English example because the final vowel is realised so as to satisfy harmony; and the influencing distinctive features [-back, +high] are not adjacent to the influenced element. A morpho-phonological rule can be stated thus: the high back vowel [u] undergoes fronting to [i] in the clitic 3rd person possessive and object pronouns {-hu}, {-humaa}, {-hum} and {hunna} when these are preceded by the high front vowel [i] or by the semivowel [j], both being characterized by the phonetic distinctive features [-back, +high]. The long back open vowel of the singular feminine pronoun {-haa}, on the other hand is not subjected to the same influence, most probably for the sake of preserving gender distinction.

A similar phenomenon of vowel influence occurs with the 1st pers. possessive pronoun suffix {-ii} when attached to nouns or to a number of prepositions such as {maʕa}, {qabla} and {baʕda}, ‘with’, ‘before’ and ‘after’. But not only is the influence here regressive, for the underlying phrase /maʕa+ii/, ‘with me’ is realised [maʕii] and /kitaabu+ii/→[kitaabii], ‘my book’, but it also results in the overwriting and deletion of the final vowel to

which the pronoun {-ii} is attached, and therefore to resyllabication, in order to avoid the unacceptable and prosodically-undefinable syllabic structure – CVVV]⁸, and to satisfy the final open heavy syllable – CVV].

These and other instances can be used as evidence for the tight relation between morphological structures in Arabic and prosodic features or the syllable make-up, on the one hand, and syntactic/semantic structures, on the other. Bird and Blackburn (ibid.) clearly illustrate these relations with reference to the morpheme, for example, which, as they say, “may be represented as a complex consisting of a semantic expression, a constraint on morpho-syntactic distribution, and a phonological description”.

The model referred to as Optimality Theory, and developed by Prince and Smolensky (1993/2004), attempts to explain interrelated grammar structures in terms of constraint influences on surface forms or outputs resulting from various phonological, morphological and syntactic processes. Because of its highly complex and rich structures, the Arabic language has been regarded as a fascinating candidate for theoretical advances. One process that is relevant to our study has been pointed out by McCarthy, 2007:263, who writes:

⁸ It might be worth noting that Algerian Arabic [mʕaja] corroborates this view: there is no need to drop the vowel in question as the {-ii} pronoun here is not realised as a long vowel but as {-ja} and thus the syllable –cv] is satisfied. The same occurs with the noun phrase [χuja], ‘my brother’, in the dialect as opposed to CA/MSA [ʔaxii].

In Arabic or German, every syllable must have an onset, *tout court*, and [ʔ] is epenthesis whenever it is needed to ensure that outcome.

3.2 Vowel Epenthesis

In this kind of morphology/phonology interface, Arabic has recourse to vowel insertion before initial consonant clusters to satisfy the onset constraint, as its words cannot begin with two or more consonants.

In the imperative mood, for example, the stem /ktub/ can only be rendered by means of vowel epenthesis giving [uktub]; but as Arabic does not allow onset-less syllables, a further epenthesis, that of a glottal stop (*hamzat u lward*), is necessary to have the ‘optimal’ output [ʔuktub].

Similarly, citing the example *alqalam* (the pen), for which the article *al* necessitates a glottal epenthesis after a pause in Arabic to obtain the output [ʔalqalam], Prince and Smolensky (1993:24) acknowledge that “Arabic syllables must be of the form CV(X); epenthesis ensures that the obligatory onset is present, whatever the input configuration.”

If the onset constraint is raised, that is, when the items in question are not found in initial position, then *hamzatu lward*, the [ʔ]-epenthesis, does not apply, as in [ʔxuði ʔlqalam(a) ʔumma ʔktub ʔismak(a)] ‘Take the pen, then write your name’; (the symbol ʔ is used here to represent the juncture, the position at which [ʔ]-drop occurs or a whole [ʔ+vowel] haplology in the case of the imperative). What is noteworthy here is the interference between the

phonological process and prosodic featuring as it leads to an overall resyllabification in the utterance: [ʕu . ðil ʔa . la . m(a) ʔum . mak . tu . ʔbis . mak]. Another glottal stop, called *hamzatu l qaṣ*, does not undergo the same process because of its status as part of the root or the stem as in the causative verb /ʔadʕal/ ‘to cause to enter’, or in the broken plural template ʔafʕaal as in /ʔaṭʕaal/ ‘children’. This type of glottal stop in such cases is realised in all phonetic environments; compare [wadʕul], ‘and come in’, with elision of the epenthetic [ʔu] with [wa ʔadʕili l ʔaṭʕaal], ‘and let the children in’, where [ʔ] is not epenthetic, but part of the stems.

A phonological rule related to the definite article {al-} in Arabic concerns the total assimilation of /l/ to the next consonant if it is one of the 13 coronal obstruents⁹. But such geminating rule only applies with the article, as in /al ʕams/→[aʕʕams] ‘the sun’, or /al ramz/→[arramz]¹⁰, ‘the symbol’. Indeed, the /l/ in any other phonological environment is not submitted to the same type of assimilation, as in [qul sala:m] ‘Say peace’ or in [talsafu] ‘It stings’. This clearly shows how morphological/phonological interface can be restricted to certain contexts, and thus, Arabic computational linguistic programmes and machine translation models must be

⁹ In Arabic, the 13 coronal consonants, with /l/, are called *alkuruuf alʕamsiyya* ‘the solar letters’ because of this type of assimilation, as opposed to *alkuruuf alqamariyya*, ‘the lunar letters’, before which /l/ of *al-* keeps its lateral realisation, as in [alqamar] ‘the moon’.

¹⁰ The letter *laam* ‘ل’, however, is kept in the orthography of words like الشمس or الرّمز and gemination is represented by the *ḡaddah* diacritic symbol.

provided with in-depth analyses of the various aspects of the structure of the language and the relations between its linguistic levels as well as processes that do not conform to overall rules.

3.3 Weak Verbs Morphology

We would not like to end this article without mentioning, at least concisely, particular verb forms which do not comply with the general verbal morphology of Arabic: these are called ‘weak verbs’ because of the presence of *w* or *y* in the roots. These ‘weak’ consonants are responsible for verbal paradigms whose surface forms escape the canonical patterns of ‘strong’ verbs. A hollow verb (in which the middle radical is a weak consonant) like *qaal* ‘to say’, for instance, surfaces with a long vowel [aa] while the underlying *faʕal* pattern is /qawal/. What is so complex, but interesting at the same time, is that the surface forms of weak verbs appear to give ‘value’ to rules of easier realisation: we assume indeed that it is easier for a speaker to say [jaquulu], ‘He says’, in the imperfective than to produce the underlying /jaqwulu/; and [qul] in the imperative is much less heavy than the canonical pattern /ʔuqwul/. Similar phonological processes and other irregularities occur with other types of weak verbs, including those having a *hamza* in the root as in /ʔakal/ ‘to eat’, or a gemination, as in /madd/, ‘to extend’, whose underlying form is /madad/.

There have been various attempts at explaining all such phenomena in terms of different models: ‘rewrite rules’, prosodic morphology, optimality theory, etc. But one important thing to bear in mind is that only eclectic considerations of a language can reveal the complex and

subtle relations between linguistic items at different levels.

Conclusion

The instances we have mentioned in this paper show the clear mutual influence of phonological and morphological with prosodic elements and even syntax.

While some models, particularly Koskeniemi's 1983 two-level morphology, have been satisfactory for the description of linear language morphologies in which the concatenation of morphemes is rather simple, as in *un+want+ed* for instance, such non-linear attributes as infixation and templatic morphology found in Arabic require other models of analysis. The theory of Prosodic Morphology, proposed in McCarthy and Prince (1986), puts forward suggestions that attempt to clarify the nature of phonological structure and its effects on morphology. Their work has become a standard morphological model in computational linguistics.

A computational dictionary would not be able to recognize or to generate morphological alternations of a given word unless it is programmed on the basis of the morphological patterns of a language, as well as its phonological rules in the case of a programme supplied with a speech recognition system. A language processor would not 'understand' the English utterance [wil 'miɫjə] 'We'll miss you' or the Arabic plural [kutub] 'books' if it does not 'know' the phonological rules that involve assimilation and sound deletion in English, and the pattern system of pluralisation in Arabic, particularly its so-called 'broken' plural forms.

The point is that phonological and morphological processes in Arabic are so intimately intertwined for the generation of the various morphological distinctions and categories.

Therefore, a faithful morphological parser will have to be devised so as to be responsible for the implementation of all the rules applicable in language X, which is by no means unproblematic as the structural levels of the language are often intertwined and characterized by mutual interactions, particularly in morphologically highly complex languages like Arabic. As the fields of computer science and linguistics have grown up together during the past several decades, they have each benefited from cross-fertilization.

References

- BIRD, S. and BLACKBURN, P. 1994. 'A Logical Approach to Arabic Phonology'.
- KIRAZ, G.A. 1996; 'Computing Prosodic Morphology'. University of Cambridge (St John's College)
- MCCARTHY, John. 1979. *Formal Problems in Semitic Phonology and Morphology*.
- MCCARTHY, John (ed.) (2004). *Optimality theory in phonology i a reader*. Oxford: Blackwell.
- MCCARTHY, J.J. 2007. "What is Optimality Theory?" in *Language and Linguistics Compass* 1/4 (260–291). Journal Compilation, 2007 Blackwell Publishing Ltd.

- MCCARTHY and PRINCE [1986] (revised 1996).
Prosodic Morphology 1986
- PRINCE, Alan, and Paul SMOLENSKY. 1993/2004.
Optimality Theory: Constraint Interaction in Generative Grammar. London: Blackwell.
- SOUIDI, A. , CAVALLI-SFORZA, V. and JAMARI, A. 2003; *A Computational Lexeme-Based Treatment of Arabic Morphology*.
- TEEPLE, D. 2005. 'Intra-Paradigmatic Contrast in Arabic Verbal Morphology', from his MA thesis, 2003.