# Chapter 2
# Is the Arabic Mental Lexicon Morpheme-Based or Stem-Based? Implications for Spoken and Written Word Recognition

**Sami Boudelaa**

**Abstract** There are two contending views of Arabic morphology. The first is a morpheme-based approach which holds that Arabic surface forms consist of a root and a word pattern. The second is a stem-based approach which dispenses with roots and word patterns and views the Arabic lexicon as being built around processes that take the stem as a basic unit. The two views have implications for the way Arabic words are accessed and stored in the mental lexicon, for the patterns of deficits seen following brain injury, and for the way in which language processing is neurally instantiated in the brain. In this chapter, the different predictions of the two views are evaluated, and an *obligatory morphological decomposition* (OMD) model is suggested and compared to a dual route account and a connectionist account. The OMD is found to be superior and it is concluded that the Arabic, and indeed the Semitic lexicon, are organized in terms of morphemes which govern spoken and written word recognition processes.

**Keywords** Arabic morphology · Stem-based lexical acces · Spoken and written word recognition · Root-based lexical access

## 2.1 Introduction

Our long-term knowledge of language requires us to store information about the words of our language—what they sound like (phonology), what they mean (semantics), and how they are combined to construct utterances (syntax). This array of information is made available to us when we hear or see words. When native

---

The research reported in this article was supported in part by a United Arab Emirates University grant (FHSS222011) to Sami Boudelaa, by the U.K. Medical Research Council (U.1055.04.002.00001.01) and by the European Research Council Neurolex Advanced Grant to William Marslen-Wilson.

---

S. Boudelaa (✉)
Department of Linguistics, United Arab Emirates University,
15551 Al Ain, United Arab Emirates
e-mail: s.boudelaa@uaeu.ac.ae

Neuroscience of Language Laboratory, NYU Abu Dhabi, Abu Dhabi, United Arab Emirates

E. Saiegh-Haddad, R. M. Joshi (eds.), *Handbook of Arabic Literacy,* Literacy Studies 9,
DOI 10.1007/978-94-017-8545-7_2, 

speakers of English, for example, hear the word *walk*, they will recognize it as a phonological form they have encountered before, and will recover both its meaning and its syntactic characteristics (as a verb or a noun), where *walk* as a noun behaves differently from *walk* as a verb. But many English surface forms—and even more so in other languages—are more complex than the simple form *walk*, as in the inflectional variants *walks, walking, walked*. The question therefore arises about the kind of representation used and accessed when such complex forms are read. Does each of these complex forms have its own stored representation, or are they analyzed into their stems (e.g., *walk*) and their respective suffixes (e.g., ~s, ~ing and ~ed)? Similar questions arise with regard to the representation of complex derivational forms, such as *walker* or *walkable*, as well as the issue of whether semantic factors play a role in the choice of decomposed or full-form representations.

These questions become particularly relevant when considering languages like Arabic where at least two views contend to account for how complex forms are represented in the lexicon and processed on-line. The first is the traditional *root and pattern* view, originating with the medieval Arabic lexicographers, taken on board with some important modifications by structuralist linguists (Cantineau, 1950a, b; Cohen 1961; Cohen 1951), and ultimately formalized in terms of autosegmental phonology (McCarthy 1979, 1981, 1982). The second is the *stem-based or word-based view*, motivated by certain theoretical hurdles faced by the root and pattern model on the one hand, and on the other by the pressure felt by theorists to account for Semitic morphology using the same set of universal constraints or rules applied across the world languages (Benmamoun 1998, 2003; Gafos 2003; Heath 1987, 1997, 2003; Ratcliffe 1998, 2004).

### 2.1.1 The Root and Pattern Model

There are at least three distinct versions of the root and pattern model which differ either in terms of the number of morphemic units they posit and/or in terms of the way surface word forms are thought to be built. According to the oldest version of the model which we owe to the medieval Arab grammarians, the workings of the morphological system hinge on two morphemes: a consonantal root which conveys a broad semantic meaning, and a vocalic word pattern which conveys non-referential aspects of meaning such as *perfective*, or *active*. These two units are interleaved to build a deverbal noun stem (called *maṣdar*).[1] However, the derivation of all the other surface forms does not involve root and pattern combination, but proceeds on the bases of the *maṣdar* using different morpho-phonological procedures such as prefixation, infixation, and vowel deletion or insertion. To illustrate, the root QTL 'kill' for example is initially combined with the pattern *faʕl* (CaCC) to form the stem *QaTL* 'killing'. However, to build the *place noun maQTaL*, the root QTL is not

[1] Note that this is the view of the grammarians of Basra, which is different from the view held by those of Kufa, who viewed the perfective verb form as the starting point in the derivational chain.

mapped onto the pattern maCCaC; instead the perfective verb *QaTaL* 'he killed' is derived from the stem *QaTL* 'killing' by vowel stem modification. The imperfective form *yaQTuL* is then derived from the perfective *QaTaL*, and is in turn used to derive the form *maQTaL* (Bohas and Guillaume 1984). Thus, the relationship between the root QTL and a complex form like *maQTaL* is remote and mediated by three surface forms. In contrast to this, the second version of the root and pattern view, put forward by structuralistlinguists such as Cantineau (1950a, b) and Cohen (1951), and advocated more recently by lexicographers like Hilaal (1990), conceives of every surface form as a combination of a root and a word pattern, and of the lexicon as a repository of roots and patterns with a set of rules to associate them. On this view, the difference between *QaTL* and *maQTaL* is that the root QTL is mapped onto the pattern CaCC in the first, but onto maCCaC in the second.

The third instantiation of the root and pattern model is developed within the framework of autosegmental phonology (McCarthy 1979, 1981, 1982). Here Arabic morphology is thought to operate with three morphemes: a consonantal root still believed to convey the core semantic meaning, a vocalic melody conveying morpho-syntactic meaning such as active-passive, and a CV-Skeleton that contributes morpho-syntactic information as well as determining the phonological structure of the surface form. According to this model, a form like *KaTaM* 'remained silent' is comprised of the root KTM, the vocalic melody *a*, and the CV-Skeleton CVCVC. A more complex form like *maKTaB* 'office' is analyzed into a locative prefix ma~, a consonantal root KTB, a vocalic melody *a*, and a CV-Skeleton CVCCVCVC. So much like the structuralist approach, McCarthy's model entails that the root, the vocalic melody and the CV-Skeleton are combined to derive every surface form, although in later developments of his theory some word formation processes like broken plural and diminutive formation do not operate on roots, CV-Skeletons, and vocalic melodies, but on prosodically defined portion of the input (McCarthy and Prince 1990).

Despite the differences between the three versions of the root and pattern model, there is an interesting unity underlying their apparent diversity. Specifically, they all assign a morphemic status to the root and the pattern, whether the latter is viewed as a unitary construct, as in the Arab grammarians' and the structuralists' approaches, or as a composite construct consisting of a vocalic melody and a CV-Skeleton as in McCarthy's approach.The morphemic status of roots and patterns hinges on two kinds of argument. The first is distributional, based on the observation that consonantal roots and word patterns appear in many words with overlapping meanings. For example, the root KTB surfaces in 31 forms all of which, save one form *KaTi:Bah* 'squadron', revolve around the general meaning of *writing* inherent in this root, while the word pattern maCCaC appears in hundreds of surface forms, most of which are place nouns (e.g., *maDXaL* 'inlet', *maKTaB* 'office', *maSBaH̄* 'swimming pool'. A second type of argument derives from the patterning of certain co-occurrence restrictions, which apply to the consonants of the root but not to affixal consonants. For example, the first and second consonant of the root are generally neither identical nor homorganic such that roots like *SSM or *BMS are very rare (Frisch and Adnan Zawaydeh 2001). The same constraint applies, if less stringently, to first and third consonants with very few roots like *KBK or *MTM. The second and third consonant can be identical (e.g. MDD), but not homorganic (e.g., *MDT). Since these co-occurrence restrictions can be stated in terms of the root consonants, while intervening vowels and affixal consonants are ignored, this argues for a level of representation at which the consonantal root functions as an independent entity (Greenberg 1950; McCarthy 1981).

The morphemic status assigned to roots and patterns sets this class of models in sharp contrast with the more recent stem-based approach. We will not be juggling with the three versions of the root and pattern approach in the remaining part of this paper. Instead we will use the term *root and pattern model* to encompass only the structuralist version (Cantineau 1950a, b; Cohen 1961; Cohen 1951) and McCarthy's version (McCarthy 1979, 1981, 1982). We do this for a number of reasons: first because these two versions suffer from similar problems relating to the derivation of certain forms such as the broken plurals, diminutive nouns, or place nouns, which have either directly (McCarthy's version) or indirectly (structuralist version) motivated the emergence of stem-based or word-based accounts. Second, the most significant difference between the structuralists' view and McCarthy's view relates to McCarthy's fractionation of the word pattern into a CV-Skeleton and vocalic melody. However, in previous psycholinguistic research we have found only partial evidence for the parsing of the word pattern into further components (Boudelaa and Marslen-Wilson 2004). Therefore, we take the two versions to be cognitively equivalent. Third, both versions provide similar predictions regarding the cognitive relevance of the root and the pattern and these can be clearly pitted against those derived from a stem-based or word-based approach.

### 2.1.2 The Stem-Based/Word-Based Model

Like the root and pattern approach, the stem-based model is not a homogenous approach, but has a number of different versions. For instance, Heath (1987, 1997, 2003) draws a distinction between lexical representations, morphological derivation, and lexical processing and argues, on the basis of observations such as the above, that as far as lexical representation and morphological derivation are concerned, the consonantal root is best "consigned to oblivion" (Heath 2003, p. 115). There is no principled way, according to Heath (2003), to segregate consonants and vowels and assign them to different levels of representation. This is because the word patterns, or ablaut templates as he refers to them, cannot be said to contribute any grammatical information in many cases. Stems such as *XuBZ* 'bread', *KaLB* 'dog' and *SiLM* 'peace' abound in the language, yet their respective word patterns CuCC, CaCC and CiCC do not convey any grammatical information. On this view the stem is taken to be the singular form for nouns (e.g., *KaLB* 'dog', *QaMaR* 'moon', *Ba:B* 'door') and the imperfective form for verbs (e.g., *KTuB, XRuJ*). Where lexical processing is concerned however, Heath acknowledges that *"root-like strings are extracted from input representations […] but these extracted consonantal sequences do not correspond exactly to the traditionally recognized roots, particularly where vowels and* *semi-vowels are concerned"* (Heath 2003, p. 126–128). Accordingly he speculates that an input like *KaWwiN* 'bring into being, causative' is initially decomposed into the causative word pattern CaCCıC and the root KWN; subsequently this root is identifiedas the stem *KuN* 'be, imperfective', based on the fact that medial geminated /w/ in causative is usually the result of mapping a vowel /*u*/ onto a CC slot (Heath 2003, p. 128).

Other stem-based accounts dispense with roots and patterns altogether (Benmamoun 1998, 2003; Ratcliffe 1998, 2004). For instance, Ratcliffe (2004) suggests a sonority-based mechanism serving to strip off affixes and recover the stem, and although he is not explicit about the tripartite distinction between derivation, lexical representation and processing, his analysis carries unmistakable overtones that the stem is the pivotal element governing all three domains. Similarly, Benmamoun (1998, 2003) defines the primitive of Arabic word formation processes -and by extension of Arabic word processing- as the imperfective stem. He argues that a form like *muʕaLLim* 'teacher' is built not by mapping the root ʕLM onto the pattern muCaCCiC as the root and pattern theory would argue, but by appending the prefix *mu~* to the imperfective stem ʕaLLim.[2]

The emphasis on the stem as *the* unit of morphological representation and processing allies these accounts with the Generalized Template Theory (GTT), which suggests that constraints dictating minimal and maximal prosodic word length guide word formation in Semitic languages (McCarthy and Prince 1990; Ussishkin 2000, 2005). On this account word formation processes operate on existing words to derive new words and this is achieved by allowing word formation rules to adjust the structure of existing stems as necessary to produce the desired output. For instance, the Hebrew form *GiDeL* 'he raised' is thought to be derived by over writing the vowels *-a-a-* of the lexically stored stem *GaDaL* 'he grew *up*' without the root GDL ever being accessed as an independent element (Ussishkin 2005). Thus words with the CaCeC pattern are the primitive of morphological processing and representation in Hebrew, and presumably, those with the CaCaC pattern would serve as such a primitive for Arabic (Ussishkin 2005).

The foregoing paragraphs underline the heterogeneity of the stem-based approach. Not only are different stems posited by different theorists, but some of these theorists roundly reject the root and pattern as relevant morphological units (Benmamoun 2003; Ratcliffe 2004; Ussishkin 2005), while others concede a role for the root or a root-like unit in language processing (Heath 2003). In order to be able to adjudicate between the stem-based account and the root and pattern account, we will focus on the imperfective stem-based version as developed by Benmamoun (1998, 2003). There are three reasons for this: first, the imperfective stem version provides a unified treatment of morphological derivation arguing that both verbs and nouns can be built from the appropriate imperfective stem. Second, this account emphatically rules out any functions for the root or the pattern, and so stands in sharp contrast with the classic root-pattern model. Third, it shares with other

[2] The syllabic structure of the imperfective stem varies for different verb forms (or word patterns). So the imperfective stem for the first patterns is CCVC (e.g., *ya-DRuS* 'he studies'), but CVCCVC in the second pattern (e.g., *yu-DaRrıS* 'he teaches').

instantiations of the stem-based account the goal of aligning Semitic languages with the rest of the world languages. Consequently, if the imperfective-stem account is strained by psycholinguistic data, this will to a large extent apply to other instantiations of the stem view.

### *2.1.3 Empirical Questions*

Against this linguistic backdrop, the question we ask here, using experimental psycholinguistic and cognitive neuroscience techniques, concerns the nature of the lexical representation used and accessed by native Arabic speakers as they read or hear Arabic words. Is a form like *maSBaĦ* 'swimming pool' processed as a prefix ma~ and an imperfective stem SBaĦ, or as a word pattern maCCaC and a root SBH? To evaluate the predictions of these two approaches we bring to bear data from behavioral experiments, pathological data from aphasic patients, and recent imaging data using event related potentials. Because most of the available behavioral data are based on the *priming* task, we start by giving a brief description of this technique and its rationale.

**What is Priming?**

In a typical priming experiment, words are presented in pairs. The first member of the pair is called the *prime*, the second the *target*, and participants are usually instructed to make a lexical decision about the target (i.e., decide whether it is a word of the language or not). The relationship between the prime and target can be varied depending on the goals of the study. Prime and target may, for example, share morphological elements (e.g., *happiness/DARKNESS*), orthographic (e.g., *mile/MILL*) and/or phonological properties (e.g. *quay/KEY*), or simply be semantically related (e.g., *pledge/OATH*). Priming is said to occur when the timed response to the target (e.g., *DARKNESS*) is affected—either speeded up or slowed down—as a consequence of having previously encountered a related prime (e.g., *happiness*), relative to responses following an unrelated prime (e.g., *faithful*). The most common interpretation of priming is that the mental representations of the prime and target are interconnected or overlap in such a way that activating the representation of a prime word either activates the representation of the target word (Forster 1999; Neely 1991), or activates the representations of lexical or morphemic competitors.

In research investigating whether the mental lexicon is organized in terms of the phonetic word or the morpheme, two versions of the priming technique—cross-modal and masked priming—have been used extensively.

*Cross-modal priming*: here the prime and target are distinct perceptual events, with a visual target presented immediately at the offset of an auditory prime (Marslen-Wilson et al. 1994). Since the prime and target are in different sensory modalities, priming is thought to occur at the more abstract level of the lexical entry, since it is here that prime and target overlap, rather than at lower, more modality-specific

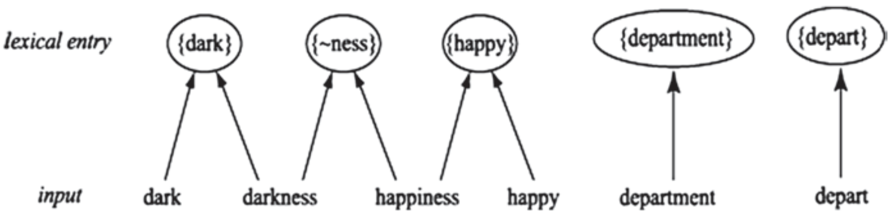


**Fig. 2.1** Illustration of how input forms sharing a morpheme map onto the same underlying representations, both for cases of a free stem as in *darkness/dark* and *happiness/happy* and a bound suffix as in *darkness/happiness*. Forms sharing a stem but no semantics map onto different representations. (Adapted from Marslen-Wilson and Zhou 1999)

levels of representation. Cross-modal morphological priming in Indo-European languages like English has been found to be contingent on the prime and target sharing either a stem and a transparent semantic link (e.g., *darkness/DARK*) or a productive affix (e.g., *happiness/DARKNESS*). Words that are related purely on historical linguistic grounds (e.g., *department/DEPART*) fail to prime (Marslen-Wilson et al. 1994). This result is theoretically captured by hypothesizing that priming between pairs like *happiness/DARKNESS* and *government/GOVERN* reflects repeated access to the same underlying lexical entry, while absence of priming among pairs of words like *department/DEPART* reflects access to different and unrelated entries, as illustrated in Fig. 2.1[3] (Marslen-Wilson and Zhou 1999).

On this view of priming, words sharing only form (e.g., *electrode/ELECT; tinsel/TIN*) do not show facilitatory priming effects, but instead compete with each other, leading to slower responses relative to the baseline. The prime *electrode*, for example, does not contain the morpheme {elect}, so hearing it as a prime will activate its own lexical entry. This will generate interference with the subsequent recognition of the target *ELECT* (Gaskell and Marslen-Wilson 2002). For pairs that are only semantically related (e.g., *pledge/OATH; curve/BEND*), cross-modal priming seems to be an elusive phenomenon (Gaskell and Marslen-Wilson 2002; Marslen-Wilson et al. 1996). It is only observed when the prime is unambiguous, and even so it decays quickly, suggesting that sharing facilitatory links between distinct lexical entries is not as effective a mechanism for priming as accessing the same shared entry.

*Masked priming*: in this version of the priming paradigm both the prime and the target are visually presented. However, participants are typically not aware that a prime is present at all, since it is presented very briefly (50 ms), and is sandwiched between a forward pattern mask (often a series of hash marks) and a backward mask, the target itself. Recent masked priming research suggests that morphologi-

[3] There are other ways of theorizing about these issues where, for example, an intermediate level of representation called the *lemma level* is thought to mediate the mapping between the input and the lexical entry (Schreuder and Baayen 1995; Taft 1994). We have evaluated these alternative views and their relevance to Arabic allomorphic variation elsewhere (Boudelaa and Marslen-Wilson 2004). Our interest here is not in comparing the contending cognitive views of morphology, but in using one of them as a starting point for predictions about how Arabic morphology may affect on-line processing.

cal decomposability alone determines priming in this task. For example, the word *corner*, which is morphologically simple but which is potentially parsable into corn + ~er, is found to facilitate the processing of the stem *CORN*, even though the two words are not in fact either morphologically or semantically related (Rastle et al. 2000). This suggests that in languages like English, masked priming does not tap into processes occurring at central levels of representation, but into early processes of word segmentation that apply to any potentially morphologically decomposable input regardless of meaning.

Cross-modal and masked priming can be seen as complementary techniques that track on-line processing at different stages, with cross-modal priming tapping into stored long-term representations, and masked priming providing a window into the early stages of lexical processing when visual input is segmented into morphemes. Should one take this view as a starting point for thinking about how the Arabic mental lexicon is organized, then a number of questions arise. First, do we see priming between Arabic words that just share a root (e.g., *maKTaB/KiTa:B* 'office'/'book')? Do words sharing a word pattern (e.g., *FaRraQ/KaSsaR* 'scatter'/'smash') also show priming? Is root priming modulated by the transparency of the semantic relationship between prime and target such that transparent pairs (e.g., *maKTaB/KiTa:B* 'office'/'book') prime, but opaque ones (e.g., *KaTi:Bah/KiTa:B* 'squadron'/'book') do not? Are the effects of priming likely to vary depending on whether we use masked or unmasked primes? As discussed below, stem-based and root-based approaches make different predictions here.

### 2.1.4 Priming Evidence for Roots and Word Patterns

*Cross-modal priming with roots*: the following illustrates a cross-modal investigation of the potential effects of the root morpheme in Arabic by using a within-word design (Table 2.1).

If Arabic roots are stored at a central level of representation and play a role similar to that played by stems in Indo-European languages, then hearing the prime word *maDXaL* 'inlet' should have two immediate processing consequences. It should activate the morpheme DXL, and at the same time inhibit other morphemes which are similar to it only in a form like DXN 'smoke', or DJL 'dupe'. The prior activation of DXL should facilitate the response to the target *DuXu:L* 'entering' when it is subsequently displayed. But what happens when the root has different interpretations across prime and target, as illustrated by *muDa:XaLah/DuXu:L* 'participation'/ 'entering' in Condition 2? Does the morpheme DXL in the prime map onto the same underlying representation as the morpheme DXL in the target in spite of their different meanings? Or are there two different entries, one for DXL meaning *participate*, and one for DXL meaning *enter*?

Having a separate entry for every meaning variation of a root, although potentially costly in terms of cognitive storage, may nonetheless lead to priming via facilitatory links between the two morphemes. A more parsimonious alternative would be to

**Table 2.1** Sample stimuli used to probe for root effects in cross-modal priming. +/−R indicates whether the prime/target pair share a root or not, +/−S whether they share semantics. Unrelated refers to the baseline condition. (For further details see Boudelaa and Marslen-Wilson 2000)

| | Prime | Target |
|---|---|---|
| 1. +R+S: Sharing a root, semantically related | مدخل<br>[maDXaL]<br>inlet | دخول<br>[DuXu:L]<br>entering |
| 2. +R−S: Sharing a root, not semantically related | مداخلة<br>[muDa:XaLah]<br>participation | دخول<br>[DuXu:L]<br>entering |
| 3. −R+S: Not sharing a root, semantically related | إيلاج<br>[ʔi:La:J]<br>insertion | دخول<br>[DuXu:L]<br>entering |
| 4. Unrelated | قهوة<br>[QaHWah]<br>coffee | دخول<br>[DuXu:L]<br>entering |

posit a unique entry to all the words featuring the same root and to associate that entry with all the shades of meaning variations that the root can have. On this scenario, hearing *muDa:XaLah* ‘participation’ should, as in condition 1, activate the root DXL, and inhibit phonologically similar roots. This activation of the root by the priming word should lead to subsequent facilitation of the target *DuXu:L* ‘entering’.

In condition 3, where the prime *ʔi:La:J* ‘insertion’ and target *DuXu:L* share meaning but feature different roots, significantly weaker priming is expected. Activation of the entry for the root WLJ ‘enter’ can only affect the entry for DXL through interlexical links, and these do not support priming as effectively as when the same linguistic entity (such as the root) is shared between prime and target (Gaskell and Marslen-Wilson 2002; Marslen-Wilson et al. 1996).

This set of predictions contrasts sharply with the predictions made by a stem-based approach where the root is thought to play little or no role in morphological processes, and the lexical entry is the imperfective stem, or the full surface form itself (Benmamoun 1998, 2003). It is not clear, on a cognitive interpretation of such an analysis, how forms sharing a root can be psycholinguistically linked to each other except on the basis of possible semantic relationships, either between their full forms or between the imperfective stems underlying primes and targets. This predicts no difference in priming between Conditions 1 and 3. In both cases there should be facilitatory activation between lexical representations of the prime and the target because of their semantic similarities, but in neither case should the strength of these effects be affected by morphological structure. Priming between words sharing a root and a transparent semantic relationship (+R+S) should be of the same magnitude as priming between pairs that are only semantically related (−R+S). For Condition 2, since the meaning of the full-form *muDa:XaLah* is not related to the meaning of the target *DuXu:L*—and similarly for their imperfective stems—there is no cognitive basis for any facilitatory priming effects, although interference may be generated given the phonological similarity between them.

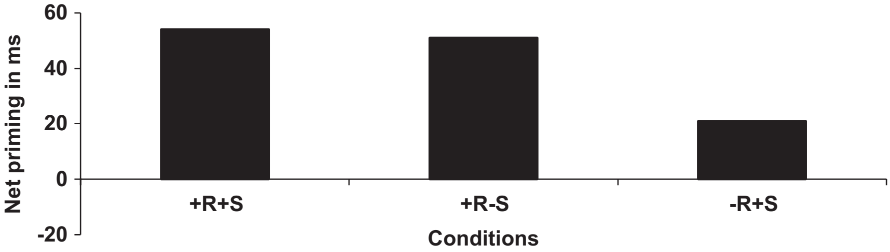


**Fig. 2.2** Cross-modal priming effects for words sharing a root with and without semantics (+R+S an +R−S), and words sharing semantics without sharing a root (−R+S)

The results, displayed in Fig. 2.2, show strong priming effects between words sharing a root. This effect is not modulated by semantic transparency. The target *DuXu:L* 'entering' is primed equally well by the semantically related *maDXaL* 'inlet', and the semantically opaque *muDa:XaLah* 'participation' (Boudelaa and Marslen-Wilson 2000), suggesting that surface forms featuring the same root map onto the same underlying lexical representation corresponding to the root morpheme. Similar results have been found in Hebrew, again using a cross-modal priming task and varying morphological and semantic relatedness (Frost et al. 2000). This finding, of overt root priming effects in the absence of a transparent semantic relationship between prime and target, seems to be specific to Semitic languages. In the Indo-European languages tested to date (e.g., English, French, Polish, Dutch), except German (Smolka et al. 2009), overt morphological priming is contingent on the transparency of the semantic relationship between prime and target, such that *casually* primes *casual*, but *casualty* does not (Marslen-Wilson et al. 1994). This cross-linguistic difference may reflect the specific structural role of Semitic morphology, especially on a root and word-pattern account, compared to a language like English, where morphological operations do not play the same fundamental role in generating the surface word-forms of the language (Boudelaa and Marslen-Wilson 2005; Marslen-Wilson 2001).

Priming effects in Condition 3, for words that are only semantically related (R+S), are much weaker and more variable. This is consistent with the view that morphological priming and semantic priming are subserved by different mechanisms: repeated access of the same underlying lexical entity in the case of morphological facilitation, and facilitatory links between different lexical entries in the case of semantic priming.

This pattern of results, which has been replicated in various forms of overt priming (including cross-modal, and auditory-auditory in Standard as well as Dialectal Arabic, where no formal teaching of roots and patterns is ever received), follows directly from a root-based account of lexical representation and processing in Semitic languages like Arabic or Hebrew. It is clearly inconsistent with a strong stem-based view on which the root morpheme would play no role in processing (Benmamoun 1998, 2003) and the only relevant unit of linguistic analysis would be the imperfective verb stem. The regularities provided by the root morpheme are picked up by the language learner and used as an organizing principle of lexical space.

**Table 2.2** Example of stimuli used to probe for word pattern effects in cross-modal priming. +WP+M stands for pairs sharing the form of the word pattern and its meaning, while +WP−M refers to pairs sharing the form of the word pattern but not its meaning. (Boudelaa and Marslen-Wilson 2000)

| | Prime | Target |
|---|---|---|
| 1. +WP+M | تجارة<br>[TiJa:Rah]<br>trade | طباعة<br>[ṭiBa:ʕah]<br>art of typography |
| 2. +WP−M | حكاية<br>[Ħ iKa:Yah]<br>story | طباعة<br>[ṭiBa:ʕah]<br>art of typography |
| 3. Phonology | مطاع<br>[muṭa:ʕ]<br>obeyed | طباعة<br>[ṭiBa:ʕah]<br>art of typography |
| 4. Unrelated | حفرة<br>[ĦuFRah]<br>hole | طباعة<br>[ṭiBa:ʕah]<br>art of typography |

*Cross-modal priming with word patterns*: similar sets of issues and contrasts arise when we turn to the second class of morphemes distinguished on a root-based approach—namely the word pattern morpheme. To probe for priming effects among words sharing a word pattern, Boudelaa and Marslen-Wilson (2000) adopted the same factorial approach as with the roots described above. The same target word was paired with four priming words as illustrated in Table 2.2.

If lexical processing in Arabic requires access to the root morpheme, as the priming data suggest, then the other component of the surface form, the word pattern, must also be accessed at some point during processing. Accordingly, upon hearing an Arabic prime word such as *TiJa:Rah* 'trade', a number of processing operations are triggered. These involve not only the activation of the root TJR 'trade', and the suppression of its cohort competitors on the one hand, but also the activation of the word pattern CiCa:Cah 'profession noun, singular' (and possibly the suppression of its cohort competitors as well) on the other. Since residual activation of the root morpheme generates priming, so should residual activation of the pattern. Therefore priming is expected in Condition 1 (+WP+M) among words sharing the phonological structure and the morpho-syntactic meaning of the word pattern as a consequence of the same underlying unit being accessed in prime and target.

By contrast, in Condition 2, the prime and the target share the phonological structure of the word pattern but not its morpho-syntactic meaning (+WP−M). The pattern CiCa:Cah has a 'profession noun' reading in the target *ṭiBa:ʕah '* art of typography', but a 'deverbal noun' reading in the prime Ħ*iKa:Yah* 'story'. If word pattern priming depends on the prime and target sharing the same morpheme, then there should be no facilitation for these pairs since their word patterns are homophonic. Alternatively, if priming through shared word patterns is much more a function of shared phonological similarity, independent of its possible linguistic interpretation, then there would be no reason not to see priming for these pairs. In Condition

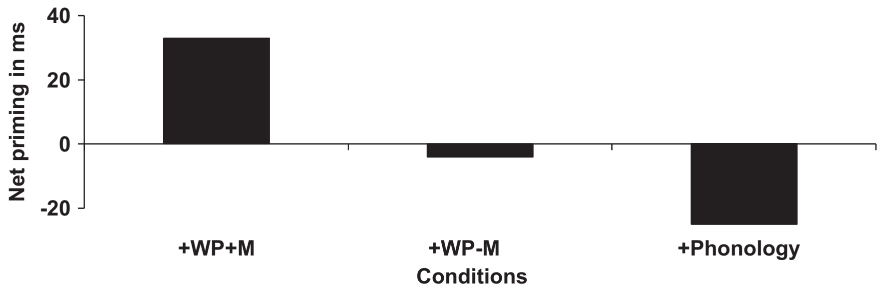


**Fig. 2.3** Net cross-modal priming effects for words sharing a word pattern (+WP), where their morphemic identity (+/−M), and words sharing phonology only (+Phonology)

3, word pairs sharing phonology like *muṭa:ʕ/ṭiBa:ʕah* 'obeyed'/ 'art of typography' should show signs of competition rather than priming.

On a stem-based approach, word patterns are not thought to be relevant linguistic units at all. So even if we relax our assumption to the point of accepting that *ṭiBa:ʕah* 'art of typography' and *TiJa:Rah* 'trade' are derived by modification of the imperfective stems *ṭ*Baʕ and Ta:JiR[4] respectively, no relationship can be established between these two items since they have different semantics and they relate to different stems. Therefore there is no basis for facilitatory priming between such items, which should behave like the phonologically related words in Condition 3 and show interference irrespective of whether they share the same morpho-syntactic interpretation of the pattern or not.

Figure 2.3 plots the net priming in the +WP+M, the +WP−M, and the + Phonology conditions. Priming between words sharing a word pattern is strongly significant, but only if the word pattern in prime and target is the same underlying morpheme, as in Condition 1 (Boudelaa and Marslen-Wilson 2000). In Condition 2, where the word patterns in prime and target are phonologically but not morphologically identical, we see no priming. This seems to be compelling evidence that the effect seen here is based on the abstract linguistic relationship between prime and target, and not on overlap in terms of their phonological or orthographic properties. Consistent with this, there is no priming in the + Phonology condition, and even some signs of interference. Taken together these results cast further doubts on the stem-based approach to Arabic in particular and Semitic in general, and provide support for the view that word pattern priming is driven by repeated activation of shared morphemes at a central level.

It is interesting to note that reliable overt word pattern priming was also found in Hebrew among verbs sharing the same word pattern (Frost et al. 2000). In Arabic, word pattern priming was found not only for verbs but also for nouns provided that the nominal word pattern occurs in the context of a productive root. Overall however, overt cross-modal priming in the two major Semitic languages, Arabic and Hebrew, provides compelling evidence for roots and word patterns as lexical units governing the process of spoken word recognition.

[4] The full orthographic forms comprising the imperfective stems *ṭ*Baʕ and Ta:JiR are respectively *yaṭBaʕu* 'he prints', and *yuTa:JiRu* 'he trades'.

**Table 2.3** Example of stimuli used to probe for word pattern and root effects in masked priming

| | Prime | Target |
|---|---|---|
| 1. +WP | شروع<br>[Š uRu:ʕ]<br>starting | دخول<br>[DuXu:L]<br>entering |
| 2. +R+S | مدخل<br>[maDXaL]<br>inlet | دخول<br>[DuXu:L]<br>entering |
| 3. +R−S | مداخلة<br>[muDa:XaLah]<br>conference | دخول<br>[DuXu:L]<br>entering |
| 4. −R + S | إيلاج<br>[ʔi:La:J]<br>insertion | دخول<br>[DuXu:L]<br>entering |
| 5. +Orthography | دخان<br>[DuXa:n]<br>smoke | دخول<br>[DuXu:L]<br>entering |
| 6. Unrelated | قهوة<br>[QaHWah]<br>coffee | دخول<br>[DuXu:L]<br>entering |

*WP* word pattern, *R* root, *S* Semantics

*Masked priming with roots and word patterns*: a growing body of visual word recognition research on Indo-European languages suggests that masked priming reflects early processes of segmentation into stems and suffixes, rather than the properties of central lexical representations (Longtin et al. 2003; Marslen-Wilson et al. 2008). This segmentation process applies automatically and blindly, so words like *corner*, which have no actual internal morphological structure, are nonetheless initially decomposed into a stem {corn} and a suffix ~er, generating masked priming to the pseudo-stem *corn*. On a root-based approach where every Arabic surface form is morphologically structured, consisting of a root and a word pattern, priming by roots and word patterns should again be found, and should not be modulated by the semantic transparency of the relationship between prime and target. We tested this prediction using the design illustrated in Table 2.3.

In Condition 1, the target *DuXu:L* 'entering' is paired with the prime *ŠuRu:ʕ* 'starting', with which it shares the nominal word pattern CuCu:C. On the traditional root and pattern approach, both *DuXu:L* and ***Š****uRu:ʕ* are morphologically structured and should be subject to the early decomposition process picked up in masked priming. This will activate the component morphemes of these words at the level of access representation and should provide a basis for priming based on repeated access to the same component CuCu:C at this level. In Conditions 2 and 3, priming is also expected based on the activation of the same morpho-orthographic component shared by the prime and target. Priming in these two conditions should be of the same magnitude although the prime and target are +R+S in Condition 2, but +R−S in Condition 3. This is because semantics does not affect the early decomposition process on which masked priming seems to be based. For the same

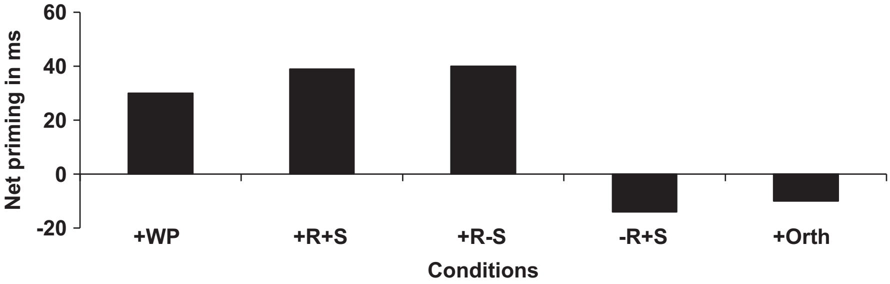


**Fig. 2.4** Masked priming effects for words sharing a word pattern (+WP), words sharing a root and semantics (+R+S), a root but no sema;ntics (+R−S), semantics but not a root (−R+S), and words sharing orthography (+Orth) at an SOA of 48 ms

reason no priming is expected in Condition 4 (−R+S), where prime and target are semantically but not morphologically related. In Condition 5, the prime and target overlap orthographically, which should lead to inhibition at the level of processing tapped onto by masked priming.

On a stem-based approach, the morphological structure provided by roots and patterns presumably has no role to play. The overlap in form between words sharing a root and words sharing a pattern should, if anything, give rise to competition between forms sharing such units at the level of access representations. This predicts that two of the three morphological conditions (+WP, +R−S) should pattern with the orthographic conditions (+Orthography), possibly showing inhibitory effects. The third morphological condition (+R+S) should pattern with the semantic condition (−R+S), although purely semantic priming is generally either weak or non-existent in masked priming at SOA's of 60 ms or less.

Figure 2.4 shows the masked priming outcome for the first five conditions relative to the Unrelated condition (Boudelaa and Marslen-Wilson 2005).

As predicted by the decompositional root and pattern model, there is a strong and statistically significant masked priming effect for both word patterns and roots (Boudelaa and Marslen-Wilson 2005). Words overlapping only in form (Condition 5: Orthography) fail to prime at all, even though they share more letters on average than words related just by a word pattern. This suggests that masked priming in Arabic hinges on the activation of potential morphemic units at the access level rather than on simple orthographic overlap between prime and target. And again there is no difference in magnitude of priming between the +R+S condition and the +R−S on the one hand and no priming in the purely semantic condition −R+S on the other. This is consistent with the claim that masked priming primarily reflects early segmentation processes—although it should not be forgotten that in Arabic, unlike English and related languages, we see no semantic interaction with morphological effects in overt priming either. Similar masked priming effects for word patterns and roots have also been reported in Hebrew (Frost et al. 1997). It is difficult to see how a stem-based approach can be modified to accommodate these findings.

### Neuropsychological Evidence

Neuropsychological research focuses on the study of damaged brain systems in order to understand normal cognitive functions. The rationale underlying research in this area is that if a specific cognitive problem can be found after an injury to a specific area of the brain, it is likely that this part of the brain in some way supports the cognitive function in question. Although the inference from patterns of breakdown to normal function is notoriously difficult and depends on the theory of normal cognition (Bullinaria and Chater 1995; Caramazza 1986; Shallice 1988), looking at the way in which the cognitive function breaks down in patients with brain damage may nonetheless be informative about the organization of the normal system. Competing morpheme-based and stem-based views of Arabic morphology seem to make different predictions about the patterns of dysfunction likely to be observed in Arabic following brain damage. Because it assigns a cognitive status to roots and patterns, the morphemic approach predicts that damage to brain areas supporting language understanding and production may affect these morphemes differentially and selectively. By contrast, on a stem-based approach neither the root nor the word pattern is referenced by cognitive processes, since the stem is the basic unit for such processes, and should be implicated in any lexically-related deficit following damage to the brain.

The Semitic neuropsychological literature offers two reports addressing the issue of whether roots and patterns can be selectively impaired. Prunet et al. (2000) assess the extent to which metathesis errors (where the ordering of elements is compromised) target root consonants as opposed to word pattern consonants. They examined the speech of an Arabic aphasic patient, ZT, suffering from stroke damage to left hemisphere territories known to be important for normal language function. When prompted to read a word like *maMLaKah* 'kingdom' from the root MLK, ZT would produce the non-word **maLMaKah* or **maKMaLah* where the order of the root consonants is swapped around, but he would almost never produce something like **KaMLamah* where a root consonant swaps positions with a word pattern consonant. A second report focused on the selective impairment of the word pattern vowels of a Hebrew speaking patient, Dudu (Barkai 1980). This patient exhibited severe problems in producing the vowels of the word pattern while his uses of the consonants of the root were preserved. For example in response to a form like *GaZaZ* 'cut', Dudu would produce the nonce form **GiZeZ*, where the order and identity of the root consonants are intact, while the vocalic pattern [a-a] is realized as [i-e].

Taken together, the cases of ZT and Dudu suggest that unless the root and the word pattern have a special cognitive status, as embodied in the morphemic approach, it would be hard to explain how errors can selectively target root consonants in ZT's case, and the vowels of the word pattern in Dudu's case (Barkai 1980; Prunet et al. 2000). Like the priming evidence reviewed earlier, the neuropsychological evidence is at variance with the stem-based approach.

**Neuro-Imaging Evidence**

Neuro-imaging techniques fall into two broad classes: haemodynamic methods, such as PET and fMRI, and electro-physiological methods such as EEG and MEG). Haemodynamic methods are predicated on the close coupling between changes in the level of activity of a neuronal population and changes in its blood supply (Raichle 1987), while electro-physiological methods take advantage of the fact that some classes of neurons act like electrical dipoles which create an electromagnetic field that can be detected and recorded from outside the head (Wood 1987). Neuro-imaging provides a means to separate and identify different cognitive operations in terms of their neuro-physiological correlates. It assumes that if two experimental conditions generate qualitatively distinct patterns of neural activity, they are more likely than not to engage functionally distinct cognitive operations (Rugg 1999).

Within this framework, Boudelaa et al. (2010) conducted an Event Related Potential (ERP) experiment to look at how the brain responds to Arabic words differing either by a consonant belonging to the root, or a vowel belonging to the word pattern. We used the Mismatch Negativity (MMN) technique which relies on electroencephalography (EEG) to measure the brain's electrical activity and make inferences about regional cortical activities (Näätänen and Alho 1997; Pulvermüller and Shtyrov 2003).

Participants were presented with two pairs of auditory stimuli while watching a silent movie. The first pair formed the root condition and was made up of the two words *ʕaRi:S/ʕaRi:F* 'bride'/ 'corporal', which feature the same word pattern CαCi:C, but different roots ʕRS and ʕRF. The second pair represented the word pattern condition and consisted of the words *ʕaRi:S*/*ʕaRu:S* 'bride'/ 'bridegroom'. These are made up of the same root ʕRS, but use the different word patterns CαCi:C and CαCu:C respectively. In both cases the word *ʕaRi:S* was used as standard and presented 85 % of the time, while the words *ʕaRi:F* and *ʕaRu:S* served as deviants in the root and pattern conditions respectively, being presented only 15 % of the time. Performance on these word-word pairs was compared to performance on closely matched non-word-non-word pairs which differed either by a consonant (e.g., *NiRi:S/*NiRi:F), or a vowel (e.g., *NiRi:S/*NiRu:S).

On a morpheme-based approach, words differing by a root consonant should elicit a different brain response than words differing by a word pattern vowel, because the diverging segment belongs to functionally distinct morphemes. The stem-based approach, in contrast, seems to predict no difference between the root and the word pattern conditions. A word like *ʕaRi:F*, represented either as a full form or in terms of an imperfective stem, is as different from *ʕaRi:S* as it is from *ʕaRu:S*.

The results are in keeping with the predictions of a morpheme-based approach, showing that at 160 ms after the deviation point, the word deviant *ʕaRi:F* 'corporal' elicits a larger MMN than its matched non-word deviant at fronto-central recording sites. There is no significant lateralization. This pattern of activation is typical of responses to content words which exhibit no inter-hemispheric differences. In the word pattern condition there was a significantly larger MMN response evoked by the word deviant ʕaRu:*S* relative to the matched non-word deviant *NiRu:S. This effect was seen in left inferior temporal regions at 250 ms after the word recognition point. These inferior temporal regions are typically associated with the processing of grammatical morphemes in earlier studies (Caplan 1992; Chapman 1999; Mohr et al. 1994; Pulvermüller 1999; Pulvermüller et al.1995). These results provide a demonstration of a neural dissociation in the processing of roots and word patterns in a Semitic language: roots, like content words in Indo-European languages, are subserved by neural assemblies equally distributed over both hemispheres, while word patterns are similar to function words, lateralizing strongly to the left. It is not clear how stem-based or full-form approaches can provide a basis for explaining these contrasts.

### *2.1.5 Implications*

The across-the-board morphemic effects described above have far reaching implications for how Arabic words are recognized from script and speech. Not only do they strongly suggest that access representations are organized in terms of roots and patterns, they also indicate that modality-free central representations of lexical form and meaning are structured in terms of the same units. Furthermore, the same units seem to govern both the early decomposition processes and the central processes of access to meaning from speech and script with the stem itself playing no role in the lexical access process *per se* (see Berent et al. 2007; Vaknin and Shimron 2011 for a different opinion). So, what kind of cognitive architecture do we need to model this? In what follows we sketch out a tentative account of morphological effects in spoken and written Arabic word recognition.We then compare this suggestion to the dual route model of Frost et al. (1997), and to the distributed connectionist account of Plaut and Gonnerman (2000); two models designed to account for similar phenomena in Hebrew.

#### **The Obligatory Morphological Decomposition Account**

The above data strongly suggest that lexical processing in Arabic evolves around roots and word patterns, and that the extraction of these units during spoken language comprehension and reading is subserved by an obligatory decomposition mechanism as schematically depicted in Fig. 2.5.

This idea is similar to Taft's (2004) and can be instantiated as an interactive activation network with localist representations corresponding to roots and word patterns. According to this view, all content words in Arabic undergo a process of *obligatory morphological decomposition* (OMD) whereby their roots and word patterns are accessed as *lexical entries*. A lexical entry will feature the morpho-syntactic, phonological, semantic and functional information associated with the component morphemes of a given word. As long as the input word has an identifiable morphological structure it will undergo decomposition whether its meaning is

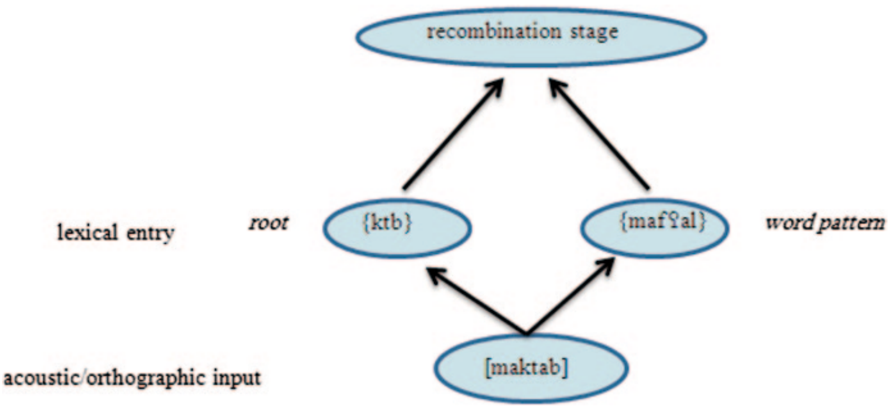


**Fig. 2.5** Schematic representation of the Obligatory Decomposition view of Arabic spoken and written word recognition

combinatorial or not. This obligatory decomposition naturally accounts for the fact that transparent forms like *Ka:TiB* 'writer' and non-transparent forms like *KaTi:Bah* 'squadron' generate comparable amounts of priming in overt (i.e., cross-modal) and covert (i.e., masked) priming.

Why would a non-transparent form like *KaTi:Bah* be decomposed when its meaning is not combinatorial? The answer to this lies in the distributional properties of the language. Most of the Arabic words sharing a root have significant overlaps in meaning. As mentioned earlier, the root KTB for instance is encountered in 31 distinct derived forms in Modern Standard Arabic, all of which, save one — namely *KaTi:Bah*- evolve around the general meaning of 'writing' (Boudelaa and Marslen-Wilson 2010). The overall consistency of the mapping between the form of a root—in this case KTB-and its semantic interpretation—'writing'-provides a valuable island of reliability in the otherwise arbitrary task of learning to relate a given spoken or written form to its meaning. This promotes the development of a parsing strategy to extract the linguistic component that helps attenuate the severity of the arbitrariness of the form-to-meaning mapping; and this unit happens to be the morpheme in Arabic.

The initial obligatory decomposition applied to non-transparent *KaTi:Bah*-like forms needs to be subsequently followed by a *recombination stage* so that their idiosyncratic interpretation can be established. So the claim here is not so much that there are no full form representations in the Arabic mental lexicon, but that such representations are available only for opaque forms, and that most importantly the storage of such forms does not circumvent obligatory decomposition. So, unlike the dual route view of Frost et al. (1997) described below, the OMD posits a single parsing route to the lexicon and a post-access recombination stage in the case of opaque forms. An interesting consequence of this is that surface form frequency effects are expected to affect processing at the recombination stage, whereas root frequency and word pattern frequency effects should affect the earlier stages of the lexical access process itself. This prediction is being tested by the present author in a lexical decision EEG experiment where variables such as surface frequency, root frequency and word patterns frequency, among others, are correlated with the Event Related Potential on a millisecond by millisecond basis.

## Frost et al.'s Dual Route Model

In Frost et al.'s (1997) model, lexical units (words) and sub-lexical units (morphemes) are both represented. Processing of Hebrew printed stimuli consists of a lexical retrieval process in which lexical units are located at the word level and a morphological parsing process in which morphemic units are extracted and located at the sub-lexical level. One of the critical features of this model is that the morphological level of representation encodes only the orthographic form -and by extension to the auditory domain only the phonology- of the root. This allows the model to account for priming among words sharing a root but an opaque semantics (e.g.,*KaTi:Bah*/ *KiTa:B* 'squadron'/ 'book').

What is not clear however is whether the morphological representation of the word pattern encodes only the phonology/orthography of this unit or whether it also encodes aspects of its morpho-syntactic meaning. If the morphological representation of a word pattern like CuCu:C or CiCa:Cah is assumed to encode only the phonological -and orthographic- attributes of this unit, by analogy to the morphological representation of the root, then the model would predict facilitation among primes and targets that share the phonological structure of a word pattern but not its morpho-syntactic meaning such as *TiJa:Rah*/QiL*a:Dah* 'trade'/ 'necklace', where the pattern CiCa:Cah has a profession noun interpretation only in the prime.This is not the case however; significant priming in nouns[5] at least is observed only when the word pattern occurs in the context of a productive root and when the prime-pattern and target-pattern have the same phonological structure and the same morpho-syntactic interpretation as in *TiJa:Rah/ṭ*iBa:ʕah 'trade'/ 'art of topography' where the pattern encodes the meaning of *profession noun* in both the prime and the target (Boudelaa and Marslen-Wilson 2011).

A similar problem for the dual route model is raised by the Hebrew cross-modal priming data which suggests that +R+S prime-target pairs show evidence of stronger priming than matched +R−S pairs (Frost et al. 2000). This suggests that semantics modulates root priming in overt tasks in Hebrew and consequently that the morphological level of representation of the root cannot be claimed to represent only the form of this unit. In addition to this, the dual route model is not clear about

[5] The situation with word patterns in verbs is a bit different, with priming occurring regardless of the interpretation of the pattern (Boudelaa and Marslen-Wilson 2012). This is arguably because there are much fewer patterns in the verb morphology domain and so deviation from the correct morpho-syntactic interpretation of the pattern can be tolerated. In contrast, the nominal word pattern space is densely populated with more than 400 nominal patterns, which precludes deviation from the specific morpho-syntactic of the pattern at hand and consequently prevents priming among patterns that do not have the same meaning.

the time taken to recognize a word via the lexical route or the sub-lexical route. Do the two routes race with each other? What variables affect the race? Or is the race between the lexical and the sub-lexical routes fixed such that processing via the two pathways is instantiated simultaneously and systematically delivers an output?

## Plaut and Gonnerman's (2000) Connectionist Network Model

Plaut and Gonnerman (2000) used a simple feed-forward network, where orthographic input is mapped onto semantic output via a set of hidden units, to demonstrate that non-semantic morphological effects are not incompatible with a distributed connectionist account. Like many other connectionist accounts (e.g., Rueckl et al. 1997; Joanisse and Seidenberg 1999; Seidenberg and Gonnerman 2000), Plaut and Gonnerman's model assumes morphology to be a characterization of the learned mapping between the surface form of words, that is their orthography or phonology, and their meanings. Since morphologically related words necessarily share form, and this is mapped onto largely overlapping aspects of meaning, the internal representation of a connectionist network should pick up on this quasi-regular mapping and treat morphological structure in a combinatorial way. Priming between morphologically related but semantically unrelated words was simulated using a set of morphologically related words varying in semantic transparency. These were embedded either in a morphologically rich or a morphologically poor language corresponding respectively to English and Hebrew. Morphological priming was found to increase with the degree of semantic overlap in both languages; and morphological priming occurred between words that share morphology without semantics in the morphologically rich language (Hebrew) but not in the morphologically impoverished language (English).

When lexical knowledge is represented distributedly, words that share parts of their spelling (e.g., *muMTɪʕ* 'enjoyable', *MuTʕ*ah 'pleasure'), and map that spelling onto similar meanings 'enjoyment/pleasure', have similar effects on some of the weights; therefore exposure to one word improves performance on the other. By contrast, words that share their spelling but map onto differential meanings (e.g. *MaTa:ʕ* 'commodities', *MuTʕ*ah 'pleasure') push the weights in competing directions, and exposure to one word does not benefit processing of another. Consequently the fact that the network exhibits priming among morphologically related semantically opaque forms is in itself a success, and a clear demonstration that the way opaque items are represented and processed depends on the overall linguistic environment to which the network is exposed. If most derivative forms featuring a particular root have similar semantic interpretations and only few of them deviate from the general semantics of the root, the semantically transparent items will ally themselves to exert a coherent influence on the opaque ones such that every member of the morphological family is represented more or less componentially. Plaut & Gonnerman's model is however at odds with the data summarized above.These data do not reveal a graded morphological effect as a function of semantic transparency. Instead, what we see is comparable amounts of root facilitation among semantically transparent pairs and semantically opaque ones.

## 2.2 Conclusion

We have evaluated the predictions of two views of Arabic morphology, a morpheme-based approach and a stem-based approach in the light of the available data from cognitive psychology, neuro psychology, and cognitive neuro-science. The data from these three areas of research converge to support the view that Arabic surface forms are cognitively represented on a morphemic basis, with entities such as roots and word patterns playing a crucial role in processes of lexical access and in the structure of lexical representation (Saiegh-Haddad 2013). The stem-based approach does not seem able to accommodate this kind of data, and is further strained by much other behavioral data, such as slips of the tongue (Berg and Abd-Al-Jawad 1996), and novel word acceptability judgments (Frisch and Adnan Zawaydeh 2001). Even on a purely formal linguistic level, this model arguably suffers from significant in adequacies (Prunet 2004; Tucker 2009). On a more general level, the stem-based approach fails to strike the right balance between the aim of accounting for Semitic languages using the same set of formal apparatus used with other languages, and the aim of capturing the specificity of each individual language.

The OMD sketched above is a root and pattern based account and provides a good fit to the existing data. It provides a better fit to the data than the dual route model (Frost et al. 1997), and the distributed connectionist model (Plaut and Gonnerman 2000). Future development of the OMD will need to be informed more significantly by neural consideration in order to build a neuro-cognitively viable account of speech and reading comprehension in Arabic in particular and Semitic in general.

**Acknowledgement** This paper is supported by UAE University grant (FHSS/22/2011). The author would like to thank William Marslen-Wilson, Elinor Saiegh-Haddad and Aisha Al Naimi for valuable comments on and help with earlier versions of this manuscript.

## References

Barkai, M. (1980). Aphasic evidence for lexical and phonological representations. *Afroasiatic Linguistics, 7/6,* 163–187.

Benmamoun, E. (1998). Arabic morphology: The central role of the imperfective. *Lingua, 108,* 175–201.

Benmamoun, E. (2003). The role of the imperfective template in Arabic morphology. In J. Shimron (Ed.), pp. 99–114. Amsterdam: John Benjamins.

Berent, I., Vaknin, V., & Marcus, G. F. (2007). Roots, stems and the universality of lexical representations: Evidence from Hebrew. *Cognition, 104,* 254–286.

Berg, T., & Abd-Al-Jawad, H. (1996). The unfolding of suprasegmental representations: A cross linguistic perspective. *Journal of Linguistics, 32,* 291–324.

Bohas, G., & Guillaume, J.-P. (1984). *Étude des théorie des grammairiens arabes*. Damas: Institut Français de Damas.

Boudelaa, S., & Marslen-Wilson, W. D. (2000). Non-concatenative morphemes in language processing: Evidence from Modern Standard Arabic. In J. McQueen & A. Cutler (Eds.), *Pro-*

*ceedings of the workshop on spoken word access processes* (Vol. 1, pp. 23–26). Nijmegen: Netherlands.

Boudelaa, S., & Marslen-Wilson, W. D. (2004). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language, 90,* 106–116.

Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes, 20,* 207–260.

Boudelaa, S., & Marslen-Wilson, W. D. (2010). ARALEX: A lexical database for Modern Standard Arabic. *Behavior Research Methods, 42,* 481–487.

Boudelaa, S., & Marslen-Wilson, W. D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes, 26,* 624–652.

Boudelaa, S., & Marslen-Wilson, W. D. (2012). Structure, form, and meaning in the mental lexicon: Evidence from Arabic. *Journal of Memory & Language* (in review).

Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y., & Marslen-Wilson, W. D. (2010). Arabic morphology in the neural language system: A mismatch negativity study. *Journal of Cognitive Neuroscience, 22,* 998–1010.

Bullinaria, J. A., & Chater, N. (1995). Connectionist modelling: Implications for cognitive neuropsychology. *Language and Cognitive Processes, 10,* 227–264.

Cantineau, J. (1950a). La notion de <schème> et son altération dans diverses langues sémitiques. *Semitica, 3,* 73–83.

Cantineau, J. (1950b). *Racines et Schèmes, Mélanges William Marçais* (pp. 119–124). Paris: Maisonneuve.

Caplan, D. (1992). *Language: Structure, processing, and disorders*. Cambridge: MIT Press.

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from analysis of patterns of impaired performance. *Brain and Cognition, 5,* 41–66.

Chapman, R. M. (1999). Function and content words evoke different brain potentials. *Behavioral and Brain Sciences, 22,* 282–284.

Cohen, M. (1951). Langues chamito- sémitiques et linguistique historique. *Scientia, 86,* 304–310.

Cohen, D. (1961). Le vocabulaire de base sémitique et le classement des dialectes du sud. *Semitica, 11,* 55–84.

Forster, K. I. (1999). The microgenesis of priming effects in lexical access. *Brain and Language, 68,* 5–15.

Frisch, S., & Adnan Zawaydeh, B. (2001). The psychological reality of OCP-place in Arabic. *Language, 77,* 91–106.

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew: A masked priming inverstigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory & Cognition, 23,* 829–856.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., & Marslen-Wilson, W. D. (2000). Morphological priming: Dissociation of phonological, semantic, and morphological factors. *Memory & Cognition, 28*(8), 1277–1288.

Frost, R., Kugler, T., & Forster, K. I. (2005). Orthographic structure versus morphological structure: Principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31,* 1293–1326.

Gafos, I. A. (2003). Greenberg's asymmetry in Arabic: A consequence of stems in paradigms. *Language, 79,* 317–355.

Gaskell, G., & Marslen-Wilson, W. D. (2002). Representation and competion in the perception of spoken words. *Cognitive Psychology, 45,* 220–566.

Greenberg, J. H. (1950). The patterning of root morphemes in Semitic. *Word, 6,* 162–181.

Heath, J. (1987). *Ablaut and ambiguity: Phonology of a Moroccan Arabic dialect*. New York: State University of New York.

Heath, J. (1997). Moroccan Arabic phonology. In A. S. Kaye & P. T. Deniels (Eds.), *Phonologies of Asia and Africa Inc Caucasus* (Vol. I., pp. 205–217). Indiana: Winona Lake.

Heath, J. (2003). Arabic derivational ablaut, processing strategies, and consonantal "roots". In J. Shimron (Ed.), *Language processing and acquisition in languages of Semitic root-based, morphology* (pp. 100–115). Amsterdam: John Benjamins.

Hilaal, Y. (1990). Deriving from roots and word patterns. *Linguistica Communicatio, 1,* 77–80, (in Arabic).
Joanisse, M., & Seidenberg, M. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Science*, 1–6.
Longtin, C. M., Segui, J., & Halle, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes, 18,* 313–334.
Marslen-Wilson, W. D. (2001). Access to lexical representations: Cross linguistic issues. *Language and Cognitive Processes, 16,* 699–708.
Marslen-Wilson, W. D., & Zhou, X.-L. (1999). Abstractness, allomorphy, and lexical architecture. *Language and Cognitive Processes, 14,* 321–352.
Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101,* 3–33.
Marslen-Wilson, W. D., Moss, H. E., & Halen, v. S. (1996). Perceptual distance and competetion in lexical access. *Journal of Experimental Psychology: Human Perception & Performance, 22,* 1376–1392.
Marslen-Wilson, W. D., Bozic, M., & Randall, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language & Cognitive Processes, 23,* 394–421.
McCarthy, J. J. (1979). *Formal problems in Semitic phonology and morphology*. Unpublished Ph.D. Dissertation, MIT.
McCarthy, J. J. (1981). A prosodic theory of non-concatenative morphology. *Linguistic Inquiry, 12,* 373–418.
McCarthy, J. J. (1982). Prosodic templates, morphemic templates, and morphemic tiers. In H. Van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (pp. 191–223). Dordrecht: Foris Publications.
McCarthy, J. J., & Prince, A. (1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory, 8,* 209–283.
Mohr, B., Pulvermüller, F., & Zaidel, E. (1994). Lexical decision after left, right and bilateral presentation of content words, function words, and non-words: Evidence from interhemispheric interaction. *Neuropsychologia, 32,* 105–124.
Näätänen, R., & Alho, K. (1997). Mismatch negativity-the measure for central sound representation accuracy. *Audiology & Neuro-Otology, 2,* 341–353.
Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale: Erlbaum.
Plaut, D., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes, 15,* 445–485.
Prunet, J. F. (2004). La formation des mots en sémitique. In P. Larcher (Ed.), Aix-en-Provence: Presses de l'Université de Provence.
Prunet, J. F., Béland, R., & Idrissi, A. (2000). The mental representation of Semitic words. *Linguistic Inquiry, 31,* 609–648.
Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22,* 253–279.
Pulvermüller, F., Lutzenberger, W., & Birbaumer, N. (1995). Electrocortical distinction of vocabulary types. *Electroencephalography and Clinical Neurophysiology, 94,* 357–370.
Pulvermüller, F., & Shtyrov, Y. (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *Neuroimage, 20,* 159–172.
Raichle, M. E. (1987). Circulatory and metbolic correlates of brain function in normal humans. In F. Plum & V. Mountcastle (Eds.), *Handbook of physiology: The nervous system* (Vol. 5, pp. 643–674). Baltimore: American Physiological Society.
Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes, 15,* 507–537.

Ratcliffe, R. R. (1998). *The "broken" plural in Arabic and comparative Semitic: Allomorphy and analogy in non-concatenative morphology*. Amsterdam: John Benjamins.
Ratcliffe, R. R. (2004). Sonority-based parsing at the margins of Arabic morphology: In response to Prunet, Beland and Idrissi (2000) and Davis and Zawaydeh (1999, 2001). *Al-[c]arabiyya, 37,* 73–95.
Rueckl, J. G., Mikolinski, M., Raveh, M., Miner, C. S., & Mars, F. (1997). Morphological priming, fragment completion, and connectionist networks. *Journal of Memory and Language, 36,* 382–405.
Rugg, M. D. (1999). Functional neuroimaging in cognitive neuroscience. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language*. Oxford: Oxford University Press.
Saiegh-Haddad, E. (2013). A tale of one letter: Morphological processing in early Arabic spelling. *Writing Systems Research, 5,* 169–188.
Schreuder, R., & Baayen, R. H. (1995). Modelling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale: Erlbaum.
Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4,* 353–361.
Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.
Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes, 9,* 271–294.
Taft, M. (2004). Morphological decomposition and the reverse base frequency effect.*The Quarterly Journal of Experimental Psychology, 57A,* 745–765.
Tucker, M. A. (2009). The root-and-prosody approach to Arabic verbs. *Qualifying Paper, UCSC*, 1–43. http://people.ucsc.edu/~matucker/papers.html.
Ussishkin, A. (2000). Root-and-Pattern morphology without roots or patterns. In M. Hirotani, A. Coetzee, N. Hall, & J.-Y. Kim (Eds.), *Proceedings of NELS 30* (pp. 655–670). GLSA, UMass Amherst.
Ussishkin, A. (2005). A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language and Linguistic Theory, 23,* 169–218.
Vaknin, V., & Shimron, J. (2011). Hebrew plural inflection: Linear processing in a Semitic language. *The Mental Lexicon, 6,* 197–244.
Wood, C. C. (1987). Generators of event related potentials. In A. M. Halliday, S. R. Butler, & R. Paul (Eds.), *A text book of clinical neurophysiology* (pp. 535–567). Chichester: Wiley.