

Verbal noun formation in Classical Arabic

Alain Kihm (CNRS, Laboratoire de Linguistique formelle, Paris)

1. Introduction

Those forms Arab grammarians call *maSaadir* ‘sources, starting points’ (sg. *maSdar*) and are called verbal nouns (VNs) here fulfil the functions of infinitives or gerunds or nominalizations in Indo-European languages. Unless they are lexicalized with more or less unpredictable meanings, they can generally be translated as ‘the fact/act of Ving or being Ved’ – since VNs may often be interpreted actively or passively (see Wright 1: 114-115; and see section 7 below). As we shall see, the ways they are formed are even more wildly diversified than is the case with broken plurals (for which see Ratcliffe 1998; Kihm 2003, 2004).

Arab grammarians devoted much valuable work to listing those forms and pointing out what correlations there are between a given form and aspects of the meaning of the verb it is related to. This notion ‘being related to’ is indeed crucial, as the two main grammar schools, the Kufis and the Basris (from Kufa and Basra in present-day Iraq), conducted a lively, and often very modern-sounding debate about what this relation means: are VNs derived from the corresponding verbs, more or less in the way (*a*) *killing* is derived from (*to*) *kill*, as the former maintained? Or is it the other way around, verbs being derived from VNs, which are themselves built directly on the consonantal root, as was the Basris’ opinion? (For a thorough exposition and instructive comments on this debate, see Bohas & Guillaume 1984:124-148.)

As will become apparent in the following, the present work endorses part of the Basris’ position, insofar as I will try to demonstrate that VNs do indeed represent the root in a sometimes quite straightforward way. Adopting this part of their view does not commit us, however, to going along with the other part, namely that verbs are derived from VNs or, more relevantly for my purposes, that *only* VNs directly proceed from roots. To the best of my knowledge, no study of Arabic VNs in a modern, post-1957 framework has been attempted so far.

The present paper proposes an account based on what I tentatively call the Root-and-Sites Hypothesis (RSH). The RSH is summarily presented in the following section, and it is tested on Classical Arabic (CA) inflection and derivation in section 3.¹ Then the bare facts of CA VN formation are presented in section 4, and the RSH is applied to them in section 5, leading to the conclusion that VNs are of two types: either they consist in the bare root, or they are formed like existing or possible broken plurals, i.e. by inserting a glide in an internal site located between the 2nd and 3rd root consonants (C₂ and C₃). In VNs of derived Forms (II to X), elements pertaining to the verbal derivation proper may be retained in addition. Yet, it cannot be said that VNs are derived from verbs in the (seemingly) straightforward way *killing* comes from *kill*. ‘Going back’ to the root is always necessary. In section 7, an attempt is made to draw a connection between VNs’ morphological make-up and their voice ambiguity and case-marking properties vis-a-vis their arguments. Finally, in the conclusion, I tackle the issue of the relationship of VN and broken plural formation with apophony and templates. An appendix briefly deals with the rare Forms XI to XV. It is not included merely for the sake of

¹ By ‘Classical Arabic’ one designates the language of the Koran and of most literature down to modern times. It is this variety of the language that is described by the Arab grammarians and in Western reference grammars such as Wright (1896-1898/1991). For more particulars about the differences between CA and Modern Standard (or Written) Arabic, see Holes (1995) and Badawi, Carter & Gully (2004). Wright’s grammar actually consists in the enlarged translation of Carl Paul Caspari’s *Arabische Grammatik* (1844-1848). It was first published in 1859. I use the two-volumes-in-one 1991 reprint of the third edition. ‘Wright 1’ in the text means Wright, 1st volume.

completeness, but also because it shows how the present framework allows one to account for these odd forms in a rather straightforward way.

2. The Root-and-Site Hypothesis

The RSH originates in Guerssel & Lowenstamm's (1990) work on CA verbal derivation, later applied to plural formation by Asfour (2001) and Kihm (2003). It is a member of the inferential-realizational family of models (see Stump 2001, Chapter 1) in the sense that it imposes no bi-uniqueness constraint on the relationship between underlying features and their overt expressions. Although it makes use of representations rather than of explicit rules, it is rather close in spirit, I think, to Stump's Paradigm Function Morphology (see Stump 2001, 2002). Moreover, the RSH is agnostic *a priori* as to whether the overall workings of grammar should be couched in derivational or representational terms. The objects it works with contain features of diverse origins, some being inherent, others a correlate of the syntactic function of the item, and it does not matter at this particular level how these features were put together. That said, the RSH appears to me to be more in tune with non-derivational, globally parallel approaches to grammar such as Autolexical Syntax (Sadock 1991), HPSG (Pollard & Sag 1994; Sag, Wasow & Bender 2003), Role and Reference Grammar (Van Valin & LaPolla 1997), and Tripartite Parallel Architecture (Jackendoff 1997). Fleshing out such affinities is a task for future research, however.

I assume two lexicons: an abstract lexicon (a-lex) and a concrete lexicon (c-lex). A-lex consists in a possibly universal collection of abstract denotational or functional features. What I call denotational features (d-features) are linguistic representations of concepts of entities, qualities and events in all possible words; functional features (f-features), on the other hand, are linguistic representations of specifications of these concepts. A-lex thus constitutes the inherent ('lexical' as opposed to compositional) semantic component of grammar, directly interfaced with the conceptual faculties of mind. D- and f-features combine in a-lex to form abstract lexical representations (ALR), the exact format of which (say., as logico-semantic formulae or otherwise) is irrelevant to my purpose.

In c-lex, a-lex elements are translated into phonology-compatible elements such as roots, affixes, and processes (e.g., apophony). C-lex may thus be viewed as the interface between a-lex and the morphophonological component. Nearly all the elements represented in c-lex ultimately come from a-lex.² As already pointed out, some may be there by virtue of the syntactic position of the represented item (e.g., structural case in nouns, person agreement in verbs), some because of inherent properties of the item (e.g., gender of nouns, aspectual class of verbs), and some following a decision of the speaker or writer (e.g., number in nouns, tense in verbs). It makes no difference at this level, however. What matters is how the elements are assembled and associated with material signals, for which I use the neutral term 'exponent' in the sense of overt representative of some d- or f-feature(s).³

The a-lex distinction of d- and f-features is to a large extent matched by the c-lex distinction of *roots* and *functional sites*. Roots can be defined from two perspectives (unsatisfactorily in either case). In terms of a / c-lex association (call it the semantic viewpoint), roots typically correspond to intensional constants with a denotation in some world(s) : e.g., $K\text{Æ}T = \mathbf{cat}'(x)$ (see, e.g., Van Valin & LaPolla 1997, Chapters 3-4). Discussion is wide open, of course, about the range of the deceiving adverb 'typically'. From

² Not all because of meaningless morphs or 'morphemes' in Aronoff's (1994) terminology.

³ I would rather eschew the terminological disputes about the term 'morpheme'. I do not object to it provided one does not insist that morphemes should always be discrete objects (see Plungjan 2000, chapter 2). Perhaps 'morph' would be a better cover-all term as suggested in Spencer (2004).

the perspective of c-lex / phonology interfacing, on the other hand, roots are those elements that often show self-standing, unaugmented phonological forms in some languages (e.g., $K\mathcal{A}ET = /k\mathcal{a}et/ = cat$, $IES = /jes/ = yes$, etc.). Again, all those hedges are matter for discussion.

Whatever perspective one adopts, the small cap, pseudo phonological notation I have just used for roots (and will not continue to use) is meant to suggest there is a connection between the present notion of root and that of lexeme (see, e.g., Aronoff 1994), insofar as roots also are abstract objects that may appear under or inside various, paradigmatically organized forms (e.g., $/f\dot{u}t/$ and $/fi:t/$, $/k\mathcal{a}et/$ and $/k\mathcal{a}et-s/$). Yet, roots are less abstract than the a-lex elements they translate.

Indeed, concrete roots (\mathfrak{R}) can be defined as ordered n-tuples of elements phonologically interpreted as consonants or vowels. Thus, to take a simple (and simplified) case, $\langle_{\mathfrak{R}} k, \mathcal{a}, t \rangle$ is an English root associated with the reference (or lexeme) CAT and realized $[k\mathcal{a}t]$ in *cat*, *cats*, and *catty*.⁴ Likewise, $\langle_{\mathfrak{R}} k, l, b \rangle$ is an Arabic root associated with the reference DOG and shared (the root and the reference) by *kalb* ‘(a) dog’, *kilaab* ‘dogs’, *kalba* ‘(a) bitch’, *kaliba* ‘to be ill with rabies, to be rabid’, etc.⁵ Graphic signs such as k, \mathcal{a}, t, l, b are mere labels at this level having as their sole function to distinguish the root from all others in c-lex. Notice that English roots include vowels, whereas Arabic roots contain no members directly interpretable as vowels. (The meaning of ‘directly’ will be clarified later on.)

The general notation for roots is therefore $\langle_{\mathfrak{R}} a, b, c, \dots \rangle$, in which \mathfrak{R} may be considered a feature having the list $\langle a, b, c, \dots \rangle$ as its value (see Sag, Wasow & Bender 2003:53). This formal object can sometimes be identified with a word-form (in the sense of Matthews 1974), sometimes not. For instance, $\langle_{\mathfrak{R}} k, \mathcal{a}, t \rangle = [k\mathcal{a}t] = cat$, while no word-form is directly paired with $\langle_{\mathfrak{R}} k, l, b \rangle$. This is actually a contingent, language-specific property, so $\langle_{\mathfrak{R}} k, \mathcal{a}, t \rangle$ should not be considered less abstract than $\langle_{\mathfrak{R}} k, l, b \rangle$, although both are concrete with respect to CAT and DOG (see fn. 5). Moreover, roots *qua* formal objects are represented as ordered sets. Insofar as these same ordered sets also are potential morphophonological objects, however, I call them *sites*, defining sites as domains where morphological activity can take place.

Non-root or functional sites can as well be formalized as sets, the elements of which are features such as Number, Case, Tense, etc. from a-lex. To the difference of roots, however, functional sites are assumed to comprise exactly one member each, namely the feature the presence of which is related (or not) to some morphological process. I will therefore use notations such as $\{\varphi \Phi\}$ for a functional (φ) site hosting (or identified by) the abstract f-feature Φ . Suppose $\Phi = \text{number (NUM)}$; it follows that number exponents such as English *-s* are not associated directly with abstract NUM, but with the functional site $\{\varphi \text{NUM}\}$ or whatever combination it is part of.⁶

Functional sites are ordered or unordered vis-à-vis root sites.⁷ Let us look at the latter option first, which I represent as in (1):

$$(1) \{w \langle_{\mathfrak{R}} a, b, c, \dots \rangle \{\Phi\}\}$$

Such a formula I call a concrete lexical representation (CLR), all elements of which may *in principle* be associated with phonological entities. I take it to be equivalent to the application

⁴ The angled brackets indicate that the enclosed elements are ordered.

⁵ For convenience, I use English as the a-lex metalanguage.

⁶ This specification is crucial when number cumulates with other features such as gender, or when it is not segmentally expressed as in Arabic broken plurals. In the following I will dispense with the φ subscript.

⁷ This is where I diverge most notably from Stump (2001). Stump does not assume any ordering between lexemes and what he calls ‘morphosyntactic properties’, more or less equivalent to my functional sites. As a consequence, his model only allows for global realizations such as $\langle \text{CAT}, \{\text{NUM}\} \rangle = \text{cats}$, whereas the present model permits segmented realizations, the utility of which will appear later on, I trust.

of an inflectional rule to a lexeme in the HPSG framework (see Sag, Wasow & Bender 2003: 250ff.), which it translates into a more transparent format, I think, from a morphological viewpoint.

The larger set in (1) is labelled W as it corresponds to what appears as a minimal word having the property of lexical opacity or atomicity (see, e.g., Williams 1981, 2004). Of the two (at least) subsets of W only $\langle_{\mathfrak{R}} \rangle$ is always realized.⁸ I assume, however, that $\langle_{\mathfrak{R}} \rangle$ is obligatorily provided with at least one functional site, be it realized or not. The reason for the assumption is that roots as such are uncategorized as nouns or verbs (giving these categories functional definitions as in Croft 2001).⁹ An abstract f-feature N or V is therefore needed to turn roots into syntactically usable items, and it will always be present even if no other is.¹⁰ We are thus led to formalize the conditions according to which functional sites are or are not realized.

Given the set-theoretic formula (1) (see Cooper 1974), no ordering relation holds between the root site $\langle_{\mathfrak{R}} a, b, c \dots \rangle$ represented as an ordered set and the functional site $\{\Phi\}$ represented as a singleton set. Considering that phonological realization centrally implies ordering (linearization), I submit the following:

Functional Site Realization Axiom (FSRA) : Only functional sites ordered with respect to the root site are realized, i.e associated with overt exponents. Unordered functional sites, although morphosyntactically and semantically active, are invisible to phonology ; they are not expressed by segmental or processual exponents.

Applied to (1) the FSRA implies that $\{\Phi\}$, be it NUM , or T , or whatever, cannot be overtly expressed. Insofar as $\{\Phi\}$ is the only functional site in (1), however, it has to be N or V as just mentioned. CA examples are given below :

(2) $\{W \langle_{\mathfrak{R}} k, l, b \rangle \{N\}\} \leftrightarrow kalb$ ‘dog’ (CA)

(3) $\{W \langle_{\mathfrak{R}} k, l, b \rangle \{V\}\} \leftrightarrow kaliba$ ‘to be(come) rabid’ (CA)

As a matter of fact, nothing in *kalb*’s overt form specifically marks it as being a noun: the three consonants realize the root, whereas the vowel is arbitrary and devoid of meaning (compare *qiTT* ‘cat’ and *xubz* ‘bread’).¹¹ Similarly in (3), *kaliba*’s ‘vowel melody’ /a, i, a/ cannot be taken as a verbhood exponent. It appears in many lexemes that are not verbs for one thing (e.g., *ma’ida* ‘table’). Moreover, only the second vowel has meaning and it is aspectual, again a feature related to but distinct from verbhood.

Not being associated with dedicated phonological material, the functional sites hosting N and V are thus represented as unordered vis-à-vis the root site in accordance with the

⁸ Obviously so since the root is the bearer of reference. It may be, however, that some grammatical words (pronouns, adpositions, etc.) have no root, but they realize bare functional sites.

⁹ I cannot expand here on the appropriateness of viewing uncategorizedness as a general property of roots across languages. It is enough for my purpose that it has to be recognized for the language I am concerned with in this paper, namely Arabic, in which roots consisting in consonants only simply *cannot* be nouns or verbs. For arguments to the effect that Modern Hebrew, although closely related to Arabic, might be different in this respect, see Bat-El 2003. Concerning the psychological reality of the consonantal root in Arabic, see Prunet, Béland & Idrissi (2000).

¹⁰ No other f-feature may be present if we assume that abstract f-features set at default values – e.g., singular for NUM – are simply not translated into concrete f-features. I shall have no use for this notion in the present study, but see Kihm (2005b) who shows its usefulness in analysing Old French declension.

¹¹ In the full form *kalbun* meaning ‘a dog’, the last vowel expresses case (here nominative), and final /n/, known as *tanwîn* or *nunation*, is supposed to express indefiniteness. Both features are related to nounness, but they are not nounness *per se*.

(8) $\{w \langle_{\mathfrak{R}} a, \langle\{\Phi b\}\rangle, c, \dots\rangle\}$

Here, the functional site encompassing b is ordered vis-à-vis a and c , which it respectively follows and precedes, but there is no ordering relation between b and Φ , so that no proper exponent can be associated with Φ which is then expressed through some modification of the encompassed member b : apophony if b is a vowel ; gemination or mutation if b is a consonant. An example of the former is the English strong preterite *took*, the CLR of which is given in (9); of the latter, the CA derived Form II verb *kassara* ‘to shatter’, analysed in (10):¹⁴

(9) PHON $\quad \quad \quad /εI/ \rightarrow /U/$
 $\quad \quad \quad \quad \quad \quad \quad \uparrow$
 C-LEX $\{w \langle_{\mathfrak{R}} t, \langle_{\varphi} \{T_{\langle \text{PAST} \rangle} \varepsilon I\}\rangle, k\}$ $\quad \quad \quad$ English *took*

(10) PHON $\quad \quad \quad /C/ \rightarrow /CC/$
 $\quad \quad \quad \quad \quad \quad \quad \uparrow$
 C-LEX $\{w \langle_{\mathfrak{R}} k, \langle_{\varphi} \{V_{\langle \text{II} \rangle} s\}\rangle, r\}$ $\quad \quad \quad$ CA *kassara* ‘to shatter’

(I assume a non-morphemic account of Germanic strong preterites as in Harris 1951 or Anderson 1992, *contra* Halle & Marantz 1993.)

3. CA morphology and the RSH

Let us briefly examine how the foregoing notions fare when applied to CA inflection and derivation. We shall first consider simple verbs (Form I), then derived verbs, then broken plurals.

3.1. Simple verbs

Although Form I, especially 3rd person singular masculine Perfective, is traditionally regarded as the simplest verb form, it is by no means morphologically simple in any absolute sense. Guerssel & Lowenstamm (1990) show that in, e.g., *kataba* ‘he wrote’, *fahima* ‘he understood’, or *jazula* ‘it is abundant’, the vowel (v_2) following the second root consonant (C_2) is a thematic vowel the timbre of which, i.e. /a/, /i/, or /u/, regularly correlates with the semantic class of the verb: stative verbs have /u/, while transitive and intransitive verbs have /a/ or /i/ depending on the affectedness of the subject (see Fleisch 1979; Guerssel & Lowenstamm 1990; Larcher 2003). The correlation was noticed long ago, yet it was the merit of Guerssel & Lowenstamm’s work to give it an explicit interpretation by assigning the status of a morphological *head* to v_2 . Headness of v_2 in turn explains why derived Forms exhibit uniform vowelism in the Perfective (cf. II *kattaba* ‘he made write’, X *istafhama* ‘he inquired’, IV ‘*ajzala* ‘he gave abundantly’), whereas Imperfective vowelism depends not on the semantics of the basic verb, but on the Form itself (cf. *yukattibu* ‘he makes write’, *yastafhimu* ‘he inquires’, *yujzilu* ‘he gives abundantly’). Indeed, verbal derivation involves another head, and no form can be double-headed. V_2 is then no longer the expression of a morphological head, but just default or idiosyncratic vowelism.

I endorse Guerssel & Lowenstamm’s account, although with some reworking. In particular, I interpret v_2 ’s headness in Form I as pointing to the presence of an internal

¹⁴ No *formal* distinction is thus made between inflection and derivation, nor does it seem necessary to make one, at this level at least.

functional site encompassing the vowel slot between C_2 and C_3 . And I assume this site is identified by V , the verbalizing f-feature, which in CA, contrary to English, comes with three different values corresponding to the three possible timbres of v_2 . In the present state of the research these values can receive no precise definition – perhaps none is available. I tentatively label them $\langle A \rangle$ (for ‘Active’), $\langle M \rangle$ (for ‘Middle’), and $\langle S \rangle$ (for ‘State’) (see Fleisch 1979; Larcher 2003).

The functional site hosting V may also include tense-aspect (T) as another site unordered vis-à-vis $\{V\}$. Tense-aspect in CA is traditionally provided with two values: Perfective (PF) and Imperfective (IPF). A better solution exploiting the logic of defaultness to the full is to assume that T 's default value in CA is PF, so T is not included in CLR's when it is so valued.¹⁵ The upshot is that the value of overt T cannot be but IPF, which makes it unnecessary to notate it. We thus ensure that V in the absence of T (i.e., with default tense-aspect) is given its basic phonological form according to its value. With T present, in contrast, i.e. with non-default, imperfective tense-aspect, this value is modified in predictable ways through apophony (see Guerssel & Lowenstamm 1990, 1996; Ségéral 1995; also see Kuryłowicz 1973, Chapter II). A form such as *labisat* ‘she dressed’ may then be assigned the following CLR :

(11) $\{w \langle_{\text{R}} l, b, \langle_{\{V_{\langle M \rangle} v\}} \rangle, s, \langle_{\{N_{\langle F \rangle}\}} \rangle\}$ CA *labisat* ‘she dressed’

It comprises three sites: the root site $\langle_{\text{R}} l, b, s \rangle$ and two functional sites, one of which hosts $V_{\langle M \rangle}$ and encompasses the vowel slot (notated as v) between C_2 (/b/) and C_3 (/s/), being thus ordered with respect to the members of the root. This is shown graphically by writing $\langle_{\{V_{\langle M \rangle} v\}} \rangle$ below the line to indicate that it does not break and somehow alter the consonantal root as internal sites may do (see [10] and below). $\langle_{\{V_{\langle M \rangle}\}} \rangle$ is then expressed through the timbre (here /i/) the encompassed vowel slot puts on in accordance with V 's particular value. T is not present as an unordered subsite since its value is default PF, as just explained.

The other functional site in (11) is ordered to the right of the root site. I will accordingly call it the right functional site (RFS). It corresponds to the /-t/ suffix usually analysed as meaning 3SG feminine. Yet, $\langle 3 \rangle$ counts as the default value for person, which is why the f-feature Person (PERS) is not present in (11), just as NUM is not, since both values $\langle 3 \rangle$ and $\langle \text{SG} \rangle$ can be retrieved by default. On the other hand, in a two-gender system like that of CA, feminine gender constitutes the non-default value of the nominalizing f-feature N (see Kihm 2005a). It therefore need be specified (as $\langle F \rangle$).¹⁶ In (11), $N_{\langle F \rangle}$ is included as an agreement element matching the N non-default value of the (unexpressed) subject. As a result, the RFS in (11) includes no more than $N_{\langle F \rangle}$, expressed as /-t/, not a cumulative exponent after all, but the marker of feminine gender only (compare *labisa* ‘he dressed’ with no gender marking).

Realization – i.e., interfacing c-lex with the phonological component – is viewed as a function f_{CV} that takes CLR's as its range and returns CV (consonants and vowels) strings (for the functional view of realization, see Raffelsiefen 1992). The fact that roots consist in ordered consonants in CA suggests that f_{CV} ought to be split into two components: f_C that returns consonants and f_V that returns morphologically significant vowels (see Kihm 2003). Function applications are then composed as the realization functions apply to *ordered* sites of the CLR. (Recall that unordered sites are invisible to phonology, hence to f_C and/or f_V .)

¹⁵ Naturally, this is a language specific property.

¹⁶ When gender is default masculine, N appears unvalued as in (2). N must be present in any event, because neither term of the noun / verb contrast can be considered default, it seems.

With *labisat* ‘she dressed’ still our example, f_C applies to the root, returning an ordered sequence of fully specified consonants, /l/, /b/, and /s/. F_V applies to the internal site, where it realizes the encompassed vowel slot as /i/ given $V_{<M>}$. F_C and f_V are composed, yielding a form where the root consonants and v_2 (the only vowel with a morphological function) are specified. Finally, f_{CV} applies to the RFS returning the /-t/ suffix and one last time to the whole polynomial expression.¹⁷ The realization formula for *labisat* associated with the CLR (11) can therefore be written as in (12):¹⁸

$$(12) \left(\left(\left(\left(\langle l, b, \langle \rangle, s, \langle \rangle \rangle \right)_{f_C \rightarrow /l.b.s./} \right) \times \left(\langle \langle V_{<M>} v \rangle \rangle_{f_V \rightarrow /i/} \right) \right)_{f_{CV} \rightarrow /l.bis./} \right) \times \left(\langle \langle N_{<F>} \rangle \rangle_{f_{CV} \rightarrow /t/} \right)_{f_{CV} \rightarrow /l.bis.t./}$$

The final yield /l.bis.t./ of (12), I call the terminal morphophonological sequence (TMS) because it contains all and only the phonological elements that fulfil a function. The vowels following C_1 /l/ and C_3 /s/ in the TMS are not specified because I assume (this of course is open to discussion) they have no functionality and are epenthetically inserted ‘later’ in the phonological component. Notice they are not inserted whenever they are not phonotactically required (cf. *labistu* ‘I dressed’, *’albisu* ‘I dress’, etc.).

3.2. Derived verb Forms

For derived verb Forms, Guerssel & Lowenstamm (1990) assume an internal site between C_1 and C_2 , which they call a derivational site (DS) This site, being a morphological head, inhibits the v_2 internal site, which is why derived Forms have non significant vowelings.¹⁹ It is identified by added semantic features such as causativity, intensification, reciprocity and so forth, whereas the site encompassing v_2 is identified by semantic features inherent to the verbalized root, as we have just seen. Apparently, both types of features cannot be expressed at the same time. Another difference between the two sites, according to the authors, is that the DS breaks the root, inserting one more CV sequence between C_1V_1 and C_2V_2 . Either this sequence has its own realization, e.g. /ta/ in Form VIII *ijtama^εa* ‘to convene’ (see [7]), or it gets one from outside: for instance, spreading the phonological matrix of C_2 onto inserted C gives rise to the gemination that characterizes Form II (cf. *kassara* ‘to shatter’ in [10]).

As shown in (10), my own formalization of the process consists in having the internal DS encompass C_2 , producing ‘consonantal ablaut’, here gemination. And, given the overall semantic complexity of CA derived Forms, I assume the token value $\langle II \rangle$ for the verbal f-feature V hosted by the site. We thus obtain the equivalent of Guerssel & Lowenstamm’s double-head constraint: a root can be verbalized only once.

The full CLR of Form II *kassarat* ‘she shattered’, is then as in (13), which lays bare the difference between CA-style languages and more overtly agglutinative languages such as English or Turkish, namely that the former commit to internal sites what the latter only entrust to external sites:

$$(13) \{w \langle_{91} k, \langle \langle V_{\langle II \rangle} s \rangle \rangle, r, \langle \langle N_{\langle F \rangle} \rangle \rangle \}$$

¹⁷ I adopt a conservative view according to which the CA vocabulary exhaustively consists in substantives (nouns and adjectives), verbs, and particles, including affixes. Only the first two have roots in their CLRs, with consonant-vowel segregation as a consequence.

¹⁸ Functional sites not in the scope of f_C are shown as empty brackets. I keep the convention of writing the v_2 site below the line.

¹⁹ The internal DS may itself be inhibited by the presence of an external head realized as a prefix. I cannot enter into these details.

This is where a framework that allows morph(eme)s to be processes as well as segments proves its value. Since $\langle\{V_{\langle II \rangle} s\}\rangle$ is affix-like, f_{CV} applies to it. It cannot return a dedicated segmental exponent for $V_{\langle II \rangle}$, however, since a segment, namely $C_2 /s/$, is already present in the site, and it is impossible to realize two segments at the same time as absence of order between $V_{\langle II \rangle}$ and $/s/$ would entail. C_2 gemination is therefore the exponent of $V_{\langle II \rangle}$. This is shown in (14), which only gives the relevant part:

$$(14) ((((((\langle k, \langle \rangle, r \rangle_{fc} \rightarrow /k.\langle \rangle.r./) \times (\langle\{V_{\langle II \rangle} s\}\rangle_{fcv} \rightarrow \text{Geminate}))_{fc} \rightarrow /k.s.s.r./) \dots_{fc} \rightarrow /k.s.s.r.t./)$$

Notice that the TMS does not contain any vowel since no vowel is morphologically functional in Form II. All vowel slots are thus filled up in phonology, e.g. by linking the default vowel $/a/$ to all accessible slots – i.e. all but the last one, which remains void in *kassarāt*). It is phonology as well that sees to it that the virtual sequence $/...asasa.../$ is actually uttered $/...assa.../$. (The phenomenon is known as *idġaam* ‘contraction’ – see Fleisch 1961:141ff.).

Other Forms mobilize a functional site ordered to the left of the root site (LFS). It is the case of Form X, e.g. *istafhama* ‘he inquired’, *yastafhimu* ‘he inquires’, where $V_{\langle X \rangle}$, one meaning of which is ‘reflexive causative’ – insofar as *istafhama* may be glossed as ‘he made himself understand’ – is hosted by the LFS and realized as $/st/$. (I leave aside the issue of whether $/st/$ is mono- or bimorphemic – see Guerssel & Lowenstamm 1990 for a discussion.) And there are also Forms where both the LFS and the internal DS are involved, e.g. Form V *tadakkara* ‘he was reminded’ (cf. II *dakkara* ‘he reminded’ and I *dakara* ‘he remembered’). (Traditional wisdom has it that Form V is derived from Form II, being the reflexive of the latter. Yet there are instances of Forms V without a corresponding Form II, e.g. *takabbara* ‘he thought himself great’ next to *kabura* ‘he is great’, and no **kabbara*. This shows the better adequacy of a non-oriented model like the present one.)

3.3. Broken plurals

The root and site hypothesis accounts for broken plural formation in a straightforward way. Very briefly, like verbal derivation, it involves an internal derivational site or DS ‘breaking’ the root (see Kihm 2003, 2004 for details).²⁰ The difference from the verbal DS is positional, as the nominal derivational site (NDS), as I will call it, is located between C_2 and C_3 .

One guest of the NDS (there may be other) is the abstract f-feature NUM. As already mentioned, NUM is not present in the CLR, and the NDS consequently not activated, when its value is default singular (see the CLR for *kalb* ‘dog’ in [2]). When the interpretation is plural, meaning more than 2, NUM is mapped into the CLR and it may activate the NDS, thereby producing a broken plural.²¹ The CLR of, e.g., *kilaab* ‘dogs’ is thus as in (15), leaving aside the RFS filled by the (possibly unrealized) case vowel and the tanwīn (see fn. 11), as well as the LFS possibly hosting the definite article $/l-/$.

$$(15) \{w \langle_{\text{R}} k, l, \langle\{\text{NUM}\}\rangle, b \rangle \{N\}\} \quad \text{CA } kilaab \text{ ‘dogs’}$$

The exponent of $\langle\{\text{NUM}\}\rangle$ is a glide whose timbre is (unpredictably) $/A/$, $/U/$, or (very rarely) $/I/$ (see Prunet 1996, 1998 for the low glide $/A/$). By definition glides can occur in consonant or vowel slots. Doing the former, as is the case in (15), yields a long vowel on the

²⁰ I assume pluralization to be a derivational process in CA, perhaps generally (see Beard 1995).

²¹ NUM may also be hosted by the RFS (‘sound’ suffixal plural). It is always in the RFS when it is valued as dual (e.g., *kalbaan* ‘two dogs’). CA thus leads us to posit a three-way contrast in terms of defaultness : no NUM = singular vs. NUM = plural vs. NUM $_{\langle 2 \rangle}$ = dual.

phonological surface, hence *kilaab* (or *njuuum* ‘stars’, the plural of *najm*); doing the latter gives rise to ‘short’, ‘non-iambic’ (see McCarthy & Prince 1990) broken plurals such as *kutub* ‘books’ (sg. *kitaab*). Crucially, broken plural formation is blind to whatever stands outside the root, in particular gender endings (cf. *madiina* / *mudun* ‘town(s)’). We thus achieve a unified treatment of broken plurals of all types that is strictly linear (contrary to McCarthy & Prince 1990) and makes no use of apophony (contrary to Asfour 2001).

Having thus rounded up and sharpened our analytical tools, we now turn to VN formation in CA, beginning with a descriptive survey of VN possible forms.

4. A descriptive survey of VN formation

Wright (1: 110-112) enumerates 44 templates for VNs associated with triconsonantal (3C) verbs in ‘basic’ Form I. I give the most frequent templates according to Wright below. (The numbers following the templates are Wright’s. I give them for easy cross-reference, but they have no theoretical significance.) Unlike broken plurals, the pairings of which with the corresponding singulars are statistical and practically unpredictable in most cases, to the notable exception of nouns with quadriconsonantal and larger roots (4C+), fairly stable correlations between VN templates and the aspectual or semantic category of the finite verb can be drawn, although uncertainty still runs strong. In accordance with the account of broken plurals sketched above, I notate surface long vowels with the capital letters for the corresponding glides /A/, /I/, and /U/.

- **CaCC** (1) : transitive **CaCaC** and **CaCiC**, e.g. *qatl* ‘killing, being killed’ (*qatala* ‘he killed’); *fahm* ‘understanding, insight’ (*fahima* ‘he understood’); *jaTf* ‘act of snatching’ (*jaTifa* ‘he snatched’).
- **CuCUC** (33) : intransitive **CaCaC**, e.g. *juluus* ‘sitting’ (*jalasa* ‘he sat’); *xuruuj* ‘going out, exit’ (*xaraja* ‘he went out’).
- **CaCaC** (2) : intransitive **CaCiC**, e.g. *farah* ‘joy’ (*fariha* ‘he rejoiced’); *marad* ‘sickness’ (*mariDa* ‘he fell sick’).
- **CaCACA(t)**, **CuCUCa(t)** (28, 34) : stative **CaCuC**, e.g. *jazaala* ‘abundance’ (*jazula* ‘it is abundant’); *saraawa* ‘generosity’ (*saruwa* ‘he is generous’); *xušuuna* ‘roughness’ (*xašuna* ‘he/it is rough’); *suhuula* ‘ease, smoothness’ (*sahula* ‘it is easy, smooth’).

Wright also mentions correlations with the inherent meanings of the verbs. For instance, **CiCAC** (26) is associated with verbs expressing, in Wright’s terms, ‘flight or refusal’ : cf. *širaad* ‘flight’ (*šarada* ‘he fled’); *niwaar* ‘fact of shunning with horror’ (*naara* ‘he shunned with horror’); *ibaa^c* ‘refusal’ (*abaa^v* ‘he refused’). **CuCAC** (27) goes with verbs of ‘sickness or ailment’ : cf. *uTaas* ‘a sneeze’ (*aTasa* ‘he sneezed’); *su^caal* ‘a cough’ (*sa^cala* ‘he coughed’); or with verbs denoting sounds: cf. *nu^caab* ‘a croak’ (*na^caba* ‘he croaked’); *Suyaah* ‘an outcry’ (*saaha* ‘he cried out’). One also finds **CaCaCAn** (21) for verbs of ‘violent or continuous motion’ : cf. *Tayaraan* ‘flight’ (*Taara* ‘he flew’), *jarayaan* ‘a run’ (*jaraa^v* ‘he ran’), *wamaDaan* ‘a flash’ (*wamaDa* ‘it flashed’); **CaCiC** (37) for verbs of ‘change of place’ : cf. *rahiil* ‘a trip’ (*rahalaa* ‘he travelled’), *wamiiD* ‘flash’ (alternate VN for *wamaDa* ‘it flashed’), or of sound: cf. *šahiiq* ‘a sob’ (*šahaqa* ‘he sobbed’); and **CiCACA(t)** (29) for verbs expressing ‘office, trade or handicraft’ : cf. *xilaafa* ‘office of successor, caliphate’ (*xalafa* ‘he succeeded’), *wilaaya* ‘governorship’ (*waliya* ‘he is in charge, in command’), *kitaaba* ‘office of secretary’ (*kataba* ‘he wrote’). No such semantic correlations, supposing them to be significant, can be detected with broken plurals.

Verbs in Form I sharing the same consonantal root but with different second (thematic or aspectual) vowels have distinct meanings. Separate VNs then go with each meaning: e.g., *faraqa* ‘he divided’ has *farq* (1), while *fariqa* ‘he is afraid’ has *faraq* (2); *jahara* ‘it is plain’ has *jahr* (1) or *jahaar* (26), *jahira* ‘he is dazzled’ has *jahar* (2), and *jahura* ‘he is loud’ has *jahaara* (29). Similarly, homonymous verbs are often provided with different VNs: e.g., *hakama* meaning ‘he judged’ has *ħukm* (6) whereas *hakama* meaning ‘he curbed (a horse)’ has *ħakm* (1). Here, a parallel can be drawn with broken plurals, which may also disambiguate various meanings of the singular as in the case of *bayt* ‘house, verse’ which has *buyuut* in the first meaning, and *’abyaat* in the second one.

VNs associated with derived Forms (Form II and beyond) are not so manifold. I list the most frequent templates below. (Underlined consonants are identical. In Forms VII, VIII, IX, and X of 3C, and III-IV of 4C verbs, initial /i/ is epenthetic – see Guerssel & Lowenstamm 1990.)

- II. **taCCIC**, e.g. *taksiir* (*kassara* ‘he shattered’); or (rare) **taCCiCa(t)**, e.g. *tajriba* ‘experiment’ (*jarraba* ‘he tried’); or (even rarer) **taCCaaC**, e.g. *tahraab* (*ħarraba* ‘he welcomed’).
- III. **muCACCaCa(t)**, e.g. *mukaataba* (*kaataba* ‘he wrote to s.o.’); or (rarer and lexicalized) **CiCAC**, e.g. *kitaab* ‘book’.
- IV. **’iCCAC**, e.g. *’ijlaas* (*’ajlasa* ‘he seated’).
- V. **taCaCCuC**, e.g. *tasallum* (*tasallama* ‘he took over, he received’); or **tiCiCCAC**, e.g. *tikillaam* (*takallama* ‘he spoke’).
- VI. **taCACuC**, e.g. *taqaatul* (*taqaataluu* ‘they fought each other’).
- VII. **nCiCAC**, e.g. *inkisaar* (*inkasara* ‘it broke’).
- VIII. **CtiCAC**, e.g. *ijtimaa^c* ‘meeting’ (*ijtama^ca* ‘he assembled’).
- IX. **CCiCAC**, e.g. *ihmiraar* (*ihmarra* ‘it became red’).
- X. **stiCCAC**, e.g. *istiħsaan* (*istaħsana* ‘he considered preferable’).

The same sparseness is observed with 4C verbs, simple or derived:

- I. **CaCCaCa(t)** or **CiCCAC**, e.g. *dahraja* or *dihraaj* (*dahraja* ‘he rolled sthg’).
- II. **taCaCCuC**, e.g. *tadahruj* (*tadahraja* ‘it rolled’).
- III. **CCinCAC**, e.g. *ixrinTaam* (*ixranTama* ‘he raised the nose, he was proud’).
- IV. **CCiCCAC**: *iqṣi^craar* (*iqṣa^carra* ‘he shuddered’).

Note that feminine VNs and all VNs associated with derived Forms take feminine sound suffixal plurals, inasmuch as their meanings allow them to pluralize: cf. *kitaaba* / *kitaabaat* ‘office(s) of secretary’, *ta^criif* / *ta^criifaat* ‘definition(s)’ (Form II *’arrafa* ‘he defined’), *isTilaah* / *isTilaahaat* ‘technical term(s)’. With the same semantic hedge, masculine VNs associated with Form I have broken plurals, and so do VNs associated with Forms II, III, and IV when lexicalized: cf. *kitaab* / *kutub* ‘book(s)’, *tasniif* / *tasaaniif* ‘literary composition(s), book(s)’ (*sanf* ~ *sinf* / *’asnaaf* ~ *sunuuf* ‘kind, specimen, category’), *’isnaad* / *’asaaniid* ‘chain(s)’ (*’asnad* ‘he leaned, he supported’).

5. A theory of VN formation

We shall now examine VN formation in the light of the theoretical developments of sections 2 and 3. As hinted at in the introduction, the basic assumption I will try to substantiate is that VNs derive from the root rather than from some form of the finite verb they are paired with.

They therefore do not belong to the verb's paradigm in the sense that, e.g., 1PL Perfective *qatalna* 'we killed' and 3SG masculine Subjunctive *yaqtala* '(that) he kill(s)' are members of the paradigm of the verb meaning 'to kill'. Rather it seems we should say that the VN *qatl* 'killing, murder' belongs to the 'archi-paradigm' that rounds up all the avatars of the $\langle_{\text{R}} \text{q, t, l} \rangle$ root, and that its formation, proceeding from the root node, runs parallel to that of the verb represented by *qatala* 'he killed'.

As we shall see presently, such claims seem to be true when dealing with VNs related to Form I verbs such as *qatala*. When we come to VNs of derived verbs, however, the relational network looks more complex. While it is still feasible and, I think, adequate to derive them from the root, yet a number of templates can be seen to entertain definite relationships with the corresponding finite Form, of which they include components. Moreover, it is not to be denied that even 'root' VNs like *qatl* show verbal properties that primitive nouns like *kalb* lack (see section 7), which would justify our drawing at least a special connection between them and the verbal paradigm. Nor should we overlook the fact that VNs may become fossilized as nouns and lose these verbal properties (cf. *kitaab* 'book', originally a Form III VN paired with *kaataba* 'he wrote to s.o.').

It does not seem, therefore, one can reach a firm conclusion in that matter of paradigm membership which seems to have lain at the bottom of the Kufa-Basra quarrel. What we need, I believe, is a more sophisticated theory of the paradigm that views it as a network rather than as an inheritance tree. Such a theory would allow us to maintain two necessary claims that are hard to conciliate in 'vertical' paradigms, namely that there is no direct morphological path between VNs and their finite verb associates, but there is a specific 'side' relation, in form and in content, between the former and the latter.

Meanwhile I will demonstrate how VN formation primarily involving the root can be formalized in a Root-and-Site framework.

5.1. Form I (non-derived) VNs

Let me first assume what may be the null hypothesis for VNs in the above-mentioned framework, namely that they result from combining a root with both V and N (also see Marantz 1997), thus formally expressing the fact that VNs are nouns with verbal properties (see section 7). I thus assign the following CLR to, e.g., *fahm* 'understanding':

(16) $\{_{\text{W}} \langle_{\text{R}} \text{f, h, m} \rangle \{_{\text{N}} \{_{\text{V}} \langle_{\text{M}} \rangle \} \} \}$ CA *fahm* 'understanding'

Fahm's template **CaCC** is one of the VN templates associated with Form I. Reviewing these templates immediately reveals a sharp divide. First, there are templates (1) and (2) represented by *fahm* 'understanding' and *faraḥ* 'joy', minimal words entirely similar to 'primitive' nouns to which no morphological operations have applied (cf. *kalb* 'dog', *jabal* 'mountain'). That is precisely what I will assume for them, namely that the realization function applies to (16) returning specified consonants for the root and nothing more. No internal site is identified – which is why none is represented in (16) – and $\{_{\text{N}} \{_{\text{V}} \langle_{\text{A}} \rangle \} \}$ is invisible because unordered vis-à-vis the root site. Neither are N and $\text{V}_{\langle \text{A} \rangle}$ mutually ordered.

Despite this, however, information about the timbre of the thematic vowel (/a/, /i/, or /u/) must be retrievable, because of the correlations between it and VN templates (e.g., that **CaCC** never goes with /u/ or 'state' <S> verbs – see above). Therefore, V in (16) keeps the (putative) <M> 'middle' value it has in *fahima* 'he understood', but this value is not associated with an exponent, apart from the VN template itself. Notice, however, there may be a source of confusion in saying that V in the VN 'keeps' the value it has in the finite verb. In fact, having the same V value in the VN and the finite verb pertains to those side paradigmatic relations I advocated in the foregoing subsection. We shall see several more

examples of this. Perhaps we should elevate the observation to the status of a principle, the exact formalization of which still lies beyond our reach:

- (17) VNs are formed by combining a root with both V and N. V has the same value in the VN as it has in the finite verb to which it is linked by a paradigmatic side relation.

Realization then proceeds as in (18) and (19) where the unordered functional site hosting N and $V_{\langle M \rangle}$ in the CLR is left empty as before to point out its phonological invisibility:

- (18) $\langle \langle f, h, m \rangle \{ \} \rangle_{fc} \rightarrow /f.h.m./$ CA *fahm* ‘understanding’
 (19) $\langle \langle f, r, h \rangle \{ \} \rangle_{fc} \rightarrow /f.r.h./$ CA *faraħ* ‘joy’

Morphology performs no more. Phonology then takes up the TMSs, filling up either the first vowel slot as in *fahm*, or the first two vowel slots as in *faraħ*. (The third possibility, i.e. identifying the second vowel slot only, is excluded in CA.) The vowels in these words are thus not lexical in the usual sense, since no phonological segments are present as such in the CLR, nor are they endowed with morphosyntactic functionality, as they do not realize the content of a site, they merely have the phonological function of turning TMSs into pronounceable words. Put differently, the signs *fahm* and *faraħ* constitute the overt form of the triple association of a CLR (itself associated with an ALR) with a realization formula with a phonological processing – where I take ‘sign’ in a Saussurean as well as an HPSG sense (see Pollard & Sag 1994). This is a central assumption in the present framework.

VNs formed from the pairing of CLR, such as (16) with realizations like (18) and (19) I therefore call ‘root’ VNs. They consist in the bare root plus one or two arbitrary vowels (cf. *‘ilm* ‘knowledge’ vs. *‘alima* ‘he knew’, *šukr* ‘thank’ vs. *šakara* ‘he thanked’, etc.).

Then, there are all other templates, which turn out to be actual or possible broken plural forms. In the first batch, compare *xuruuj* ‘exit’ and *quluub* ‘hearts’ (sg. *qalb*), *širaad* ‘flight’ and *kilaab* ‘dogs’ (sg. *kalb*), *raħiil* ‘a trip’ and *hamiir* ‘donkeys’ (sg. *himaar*), *kitaaba* ‘office of secretary’ and *jimaala* ‘camels’ (sg. *jamal*), *suhuula* ‘ease, smoothness’ and *buṣuula* ‘husbands’ (sg. *baṣl*). In the second batch, we find *wamaDaan* ‘a flash’, to be compared to *wirnaan* ‘varans’ (sg. *waral*); *su‘aal* ‘a cough’ similar to *kilaab* except for the quality of the first vowel, and to *ħukkaam* ‘judges’ (sg. *ħaakim*) except for the gemination of C_2 in the latter; *saraawa* ‘generosity’ similar to *jimaala* ‘camels’ except again for the first vowel, and to *‘adaarin* ‘virgins’ (sg. *‘adr*) or *fataawaa* ‘judicial opinions’ (sg. *fatwaan*) but for the different endings, i.e. for the different elements annexed to the root (see Kihm 2004).

I will therefore assume these VNs, which I accordingly call ‘broken’ VNs, to be formed like broken plurals, insofar as they include an NDS between C_2 and C_3 (see section 8 for possible interpretations of this similarity). They differ from broken plurals on two counts, however: (i) their CLR, include V in the same unordered functional site as in root VNs; (ii) their NDS hosts N instead of NUM, but the realization correlate is the same, namely insertion of a glide.²² The CLR and the realization formula of, e.g., *xuruuj* ‘exit’ are thus as in (20) and (21):

- (20) $\{w \langle_{\text{N}} x, r, \langle \{N\} \rangle, j \rangle \{V_{\langle A \rangle}\}\}$
 (21) $\langle \langle \langle x, r, \langle \rangle, j \rangle \rangle_{fc} \rightarrow /x.r.j./ \rangle \times \langle \langle \{N\} \rangle_{fv} \rightarrow /C_U/ \rangle \rangle_{fcv} \rightarrow /x.r.U.j./$ ²³

²² Recall that N is unvalued in masculine nouns. See below for feminine broken VNs.

²³ Read $\langle \langle \{N\} \rangle_{fv} \rightarrow /C_U/ \rangle$ as: $\langle \{N\} \rangle$ is realized as the glide /U/ in consonant position. The TMS /x.r.U.j/ is then phonologized as /xuruuj/, /uu/ = /u:/, by spreading U’s features onto the adjacent vowel slots. V_1 /u/ is epenthetical, and it gets its timbre by vowel harmony. See Kihm 2003 for why the glide is returned by *f_r*.

Recall that in Form I *xaraja* ‘he went out’, the internal site hosting $V_{\langle A \rangle}$ encompasses the v_2 slot, realized as /a/. All evidence points to the existence in CA of a global templatic constraint stated below:

(22) There can be no more than one internal functional site in a root site.

Activating the NDS between C_2 and C_3 for the purpose of VN formation therefore entails that the site hosting $V_{\langle A \rangle}$ and v_2 is, so to speak, expelled from the root site. Since $V_{\langle A \rangle}$ ’s ‘anchor’ v_2 obviously stays in place, $V_{\langle A \rangle}$ can no longer be ordered vis-à-vis the root site. It thus becomes phonologically invisible. It must be shown in the CLR, however, as explained a propos (16).

Consider now feminine VNs like *suhuula* ‘ease, smoothness’. The gender is not surprising, since putting relational (*nisba*) adjectives in the feminine is CA’s main device to form abstract nouns (cf. *’insaaniyya* ‘humanity’, *jam’iyya* ‘totality’, and so forth), thus uncovering a correlation between feminine (or non-default) gender and abstraction. (“The feminine of the relative adjective... corresponds therefore to German substantives in *heit, keit, schaft, thum*, and to English ones in *head, dom, ty*, etc.” – Wright 1:163.) I therefore assume that N hosted by the NDS in such forms has $\langle F \rangle$ as a value. However, $N_{\langle F \rangle}$ cannot be expressed in the NDS, but only insofar as it is hosted by (a subsite of) the RFS, as shown below:

(23) $\{W \langle_{\text{R}} k, l, b, \langle N_{\langle F \rangle} \rangle\}$ CA *kalba(t)* ‘bitch’²⁴

There are no internal or broken feminines, in other words.²⁵

We are thus led to assume that in *suhuula* and all feminine broken VNs, $N_{\langle F \rangle}$ is inserted twice: in the NDS where it is realized as a glide *qua* N; and in the RFS where its $\langle F \rangle$ value gives it /t/ as an exponent (see fn. 24). This is formalized below:

(24) $\{W \langle_{\text{R}} s, h, \langle \{N_{\langle F \rangle} \} \rangle, l \langle \{N_{\langle F \rangle} \} \rangle\} \{V_{\langle S \rangle} \}$ CA *suhuula(t)* ‘ease, smoothness’
 (25) $(((((\langle s, h, \langle \rangle, l \langle \rangle \rangle_{fc} \rightarrow /s.h.l./) \times ((\{N_{\langle F \rangle} \})_{fv} \rightarrow /C_U/))_{fcv} \rightarrow /s.h.U.l./) \times ((\{N_{\langle F \rangle} \})_{fcv} \rightarrow /t/))_{fcv} \rightarrow /s.h.U.l.t/)$

VNs associated with Form I thus appear to be either minimal words, i.e. phonological realizations of the root, or to be identical with existing or possible broken plural forms. In either case, we are forced to the conclusion that they are not morphologically derived from the actual Form I they are paired with, but they are built on the root, exactly as broken plurals with respect to their corresponding singulars. There is one connection, though, namely the quality of the thematic vowel of Form I, information about which must be available in the VN’s CLR. Yet, this information does not exert any specific morphological effect, it only has global relevance insofar as it partially determines VN template selection.

5.2. VNs of derived Forms II to X

Turning to derived Forms, we observe a similar situation except for the absence of root VNs: derived Form VNs are all broken. This is a significant difference, as it is related, as we shall see, to the presence of various derivational materials that extend the root. Moreover, information about which Form the derived verb pertains to has to be made available in the

²⁴ $N_{\langle F \rangle}$ ’s exponent is /-t/ (as in verbs), which deletes under conditions I cannot detail here, leaving /a/ as an apparent feminine marker. I will not normally write this /t/.

²⁵ In feminine nouns ‘by convention’ such as *šams* ‘sun’, we have to assume that $N_{\langle F \rangle}$ is unordered vis-à-vis the root site, therefore not realized, but nonetheless a trigger of agreement.

VN's CLR rather than information about the thematic vowel, not present in derived Forms. Again, this is because of the correlations between Forms and VN templates, as stated in (17).

In the following, I do not quite proceed in an orderly fashion, from Form II to Form X. Rather I start with the most easily accountable templates given all preceding assumptions, using them as a stepping stone to deal with more complex formations, and allowing myself as many detours as appear necessary for clarity of exposition.

5.2.1. Form III VNs

It is right away apparent that one template, **CiCAC**, related to Form III, is also a broken plural template: compare *kitaab* 'book' and *kilaab* 'dogs'. VN formation for *kitaab* thus proceeds as for *xuruuj* 'going out, exit' analysed in the preceding section, except that the 'expelled' site because of NDS activation hosts $V_{\langle III \rangle}$ instead of $V_{\langle A \rangle}$ and it is located between C_1 and C_2 in the related finite verb, in which its exponent is a C glide (cf. Form III *kaataba* 'he wrote to s.o.'). The CLR of *kitaab* is thus the following :

(26) $\{w \langle_{\text{R}} k, t, \langle \{N\} \rangle, b \rangle \{v_{\langle III \rangle} \} \}$ CA *kitaab* 'book' (Form III *kaataba* 'he wrote to s.o.')

True, *kitaab* has been fully nounified with the meaning 'book' instead of its old meaning 'act of writing to s.o., letter' which showed a stronger semantic connection with *kaataba*, and it forms the broken plural *kutub* accordingly. This does not render the previous analysis otiose, however, because it allows us to account for the long v_2 , an unusual feature in primitive nouns. *Kitaab* is thus morphologically still a VN, even though it almost – no book being without addressees – lost the semantic properties of one.

5.2.2. Form IV VNs

Another template, **iCCAC** related to Form IV as in '*ijlaas* 'act of seating', is almost identical to **aCCAC** broken plurals such as '*aqfaal* 'bolts' (sg. *qafal*), disregarding the quality of the first vowel. In the broken plural, /'a-/ is usually analysed as an augment of unclear origin (see Radcliffe 1998). In the VN, the initial glottal stop is certainly the same as appears in related Form IV '*ajlasa* 'he seated', the CLR of which is given below:

(27) $\{w \langle \{V_{\langle IV \rangle} \}, \langle_{\text{R}} j, l, s \rangle \rangle \}$ CA Form IV '*ajlasa* 'he seated'

$V_{\langle IV \rangle}$ fills up the LFS. Its exponent is best described as being an unspecified CV syllable, the onset of which is realized as a glottal stop by default, and the nucleus is occupied by the uniform /a/ of perfective derived verbs: compare *yujlisu* 'he seats', which proves the default character of /' / in '*ajlasa*, since the initial onset is now occupied by the 3rd person masculine singular prefix *y-*, as well as the inflectional value of the following vowel, /a/ in the Perfective vs. /u/ in the Imperfective. (For more details on this analysis, see Guerssel & Lowenstamm 1990.)

The first vowel of the VN '*ijlaas*, in contrast, is most probably epenthetic. This makes me suspect that the initial glottal stop in it is not the exponent of $V_{\langle IV \rangle}$, but a mere carry-over from the finite verb or, to put it more accurately perhaps, a material clue left by those paradigmatic side relations the existence of which was suggested above.

Given this, something important should be noted: the additional initial syllable thus created turns out to be invisible to the specific process that yields the broken VN, since the NDS realized as a C glide /A/ still follows the 'original' C_2 /l/. Therefore, the initial CV, although meaningless in the VN, is not part of the root, but it fills up the LFS, as does the exponent of $V_{\langle IV \rangle}$ it is related to in the finite verb. This means the VN forming process strictly occurs within the root site; external sites are off-limit for it. It also means that not only material, but the location of the material may be shared by the VN and the related finite verb.

As we shall see, there are other options. As a consequence, $V_{<IV>}$ in the VN is unordered with respect to the root site. The CLR of *'ijlaas* is given in (28):

(28) $\{w \langle C, v, \langle_{\text{R}} j, l, \langle \{N\} \rangle, s \rangle \rangle \{V_{<IV>}\}$ CA *'ijlaas* ‘act of seating’ (Form IV *'ajlasa*)

5.2.3. Form II VNs

Three VN templates are associated with Form II, already examined in section 3.2 : **taCCiC** (e.g. *taksiir* ‘shattering’ related to *kassara* ‘he shattered’), **taCCAC** (e.g., *tahraab* ‘welcome’ related to *harraba* ‘he welcomed’), and **taCCiCa** (e.g. *tajriba* ‘experiment’ related to *jarraba* ‘he tried’). They are like *'ajlaas* in that they include an additional initial syllable in addition to the root internal NDS. They differ from it, however, as this syllable, /ta/, is not present at all in the related finite verb, whereas at least the onset is shared in Form IV finite verbs and VNs, as we saw. I have no explanation for this fact, beyond pointing out that /ta/ is found in Form V, formally related to Form II since it consists in Form II plus /ta/ (cf. *takallama* ‘he spoke’).

In Form V, /ta/ is probably a prefix. If I am right in assuming a connection between Form II VNs and Form V verbs manifested by the sharing of initial /ta/, then we must conclude that /ta/ in *taksiir* is like /i/ in *'ijlaas*, namely a mere syllable filling up the LFS, since the NDS follows the root C_2 despite its presence.

The first two templates, **taCCiC** and **taCCAC**, differ only in the timbre of the glide. Notice that /I/, although rare, is found in broken plurals as well : cf. *kaliib*, an alternative broken plural of *kalb* ‘dog’. The CLR of *taksiir* (likewise *tahraab*) is then the following:

(29) $\{w \langle t, a, \langle_{\text{R}} k, s, \langle \{N\} \rangle, r \rangle \rangle \{V_{<II>}\}$ CA *taksiir* ‘shattering’ (Form II *kassara*)

N’s exponent is /I/ in C position in *taksiir*, /A/ in C position in *tahraab*.

The third VN template related to Form II exemplified by *tajriba* ‘experiment’ apparently illustrates the splitting of $N_{<F>}$ posited above for *suhuula* ‘ease’. In *tajriba*, however, v_2 is short. We can account for this in two ways. First, we can assume that internal $\langle \{N_{<F>}\} \rangle$ is realized as an /I/ glide occupying a vowel slot as in broken plurals like *kutub* ‘books’. The problem is that there are broken plurals in short /u/ or in short /a/ (e.g., *garfa / guraf* ‘room(s)'), but none in short /i/. If we wish to pursue the parallelism between VNs and broken plurals, then, we may prefer another account, namely that the overt form of the root, /jrib/, is a root **CiCC** VN (cf. *'ilm* ‘knowledge’) with epenthetic /i/ following C_2 rather than C_1 as in *'ilm* because of the initial syllable /ta/. This of course implies that $\langle \{N_{<F>}\} \rangle$ does not actually split, but it occurs once in the RFS, as shown in (30) :

(30) $\{w \langle t, a, \langle_{\text{R}} j, r, b, \langle \{N_{<F>}\} \rangle \rangle \{V_{<II>}\}$ CA *tajriba* ‘experiment’ (Form II *jarraba*)

VNs of derived verbs can thus be root VNs no less than those of Form I, but it does not show on the surface due to the additional material brought by derivation. I call such VNs ‘quasi-root’.

Notice that *tajriba* belongs to those VNs that can form a 4C **CaCACiC** broken plural, *tajaarib*, in addition to the sound plural *tajribaat*. Putting the semantic difference aside (*tajaarib* ‘experiments’ vs. *tajribaat* ‘actions of trying’), this possibility follows straightforwardly from our assumptions. Consider (31) to be the CLR of broken plural *tajaarib* ‘experiments’, formally identical to *janaadib* ‘locusts’ (sg. *jundub*), except for unordered $V_{<II>}$.²⁶

²⁶ N bears the <F> value because broken plurals are feminine for agreement.

(31) {_w <_{3r} t, j, <{NUM}>, r, b} {N<F> {V<II>}} CA *tajaarib* ‘experiments’ (Form II *jarraba*)

How do we arrive at this formula? As explained in Kihm (2003), broken plural formation for prefixed nouns involves reanalysis, whereby the initial consonant of the prefix is counted as a root consonant (cf. *miftaah* ‘key’, with the instrument prefix /mi-/ attached to the root *ftħ* ‘open’, and the 4C broken plural *mafaatih* ‘keys’). Although the semantic factors involved are not crystal-clear (are they ever?), reanalysis seems to result from the severance (at least to an extent) of the paradigmatic side relations between the prefixed noun and the finite verb whose root it shares.²⁷ To put it in a simpler way, *miftaah* ‘key’ is not felt as belonging or being connected to the paradigm of *fataħ* ‘he opened’. When the connection is not severed, as is the case with, e.g., *mu^callim* ‘teacher’, the active participle of *allama* ‘he taught’, reanalysis does not occur and sound (suffixal) pluralization is the only option (*mu^callimuun* ‘teachers’).

This is what happens in (31). When *tajriba* is taken with the meaning ‘experiment’ rather than ‘(a) try’, i.e. when it is not understood as being the VN directly related to *jarraba*, the initial /ta/ syllable, meaningless in any event, ceases to be the content of the LFS – that is to say, the connection with Form V is suppressed. The onset is then reanalysed, becoming the C₁ of the 4C (pseudo) root *tjrb*. The NDS is then ‘pulled’ leftwards to follow the new C₂ /j/, and it is realized as /A/ in C when filled by NUM.

Notice that unordered V<II> is still present in (31), as a reminder of the derivational ‘history’ of the form. N<F>, on the other hand, can no longer be hosted by the NDS, now activated by NUM. Therefore, N<F> does not occupy the same structural position in the singular *tajriba* and the broken plural *tajaarib*, as it becomes unordered along with V<II> (and therefore deprived of an exponent). This is in perfect agreement with the, I think, well established conclusion that broken plurals are not based on the corresponding singulars, but both are parallel formations, just as VNs and finite verbs are. In the sound plural *tajribaat*, in contrast, represented in (32), N<F> combined with NUM is ordered:

(32) {_w <t, a, <_{3r} j, r, b, <{N<F> {NUM}}>>> {V<II>}} CA *tajribaat* ‘tries’ (Form II *jarraba*)

I take /-aat/ to be a cumulative exponent realizing the unordered combination of N<F> and NUM. There is no reason to assume that /ta/ is annexed to the root – although it might well be, this makes no difference.

5.2.4. Form III VNs again

The other VN template related to Form III besides **CiCAC**, namely **muCACaCat** (e.g. *mukaataba* ‘action of writing to s.o.’), may be compared with ‘feminine’ 4C broken plurals such as *falaasifa* ‘philosophers’ (sg *faylasuuf*), under the assumption that, although *mu-* is a prefix, /m/ counts as a root C₁, just like /t/ in *tajaarib* ‘experiments’. This would explain why the NDS, realized as /A/ in a C slot, follows /k/ although this consonant is the C₁ of the related Form III verb *kaataba* ‘he wrote to s.o.’. Minor discrepancies are the timbre of the (non significant) vowel after C₃, [a] instead of [i], and the fact that such broken 4C VNs, like all VNs, are triptotes (*mukaatabatun* ‘an action of writing to s.o.’), whereas 4C broken plurals are diptotes (*falaasifatu* ‘philosophers’, *tajaaribu* ‘experiments’).²⁸

Although *mu-* is clearly not a mere carry-over as are /i/ and /ta/ in Form IV and II VNs, the fact that it probably is the same prefix as appears in derived Form participles (cf. *mukaatib* ‘writing to s.o.’, *mukaatab* ‘written to s.o.’) makes one reluctant to regard it as the realization of V<III> hosted by the LFS. Where then is V<III> in *mukaataba*?

²⁷ I am grateful to Lucia Tovená for reminding me that easy solutions and semantic analysis don’t go hand in hand.

²⁸ Triptotes distinguish three cases and take the tanwīn when indefinite. Diptotes only distinguish two cases in the singular and do not take the tanwīn. There is no difference between triptotes and diptotes in the plural.

In the finite Form III *kaataba* the VDS is internally active and realized as /A/ in C, as shown in (33):

(33) $\{w \langle_{\text{r}} k, \langle \{V_{\langle \text{III} \rangle} \} \rangle, t, b \} \}$ CA Form III *kaataba* ‘he wrote to s.o.’

This implies that N cannot fill the internal NDS unless $V_{\langle \text{III} \rangle}$ is externalized or unordered. The probable non-verbal character of /mu-/ pointed to above suggests that externalization is not an option. It therefore has to be unordering.

Not simple unordering, however: the NDS would then normally follow C_2 , that is /t/, yielding ill-formed **mukataaba*. And we would still be without a story for *mu-*’s presence since, to the contrary of /’/ and /ta/, we are unable to trace it back to a Form. Consider, however, that unordering need not only be with respect to the root site; it may also be vis-à-vis a functional site. Assume, then, that $N_{\langle \text{F} \rangle}$ (<F> since *mukaataba* is feminine) occupies the internal site between C_1 and C_2 hosting $V_{\langle \text{III} \rangle}$ in Form III in such a way that $V_{\langle \text{III} \rangle}$ is unordered with respect to it. The resulting configuration is given below in shortened form:

(34) $\{w \dots \langle_{\text{r}} k, \langle \{N_{\langle \text{F} \rangle} \} \{V_{\langle \text{III} \rangle} \} \rangle, t, b \} \dots \}$

Such a formula does not violate the templatic constraint (22) since there still is one internal site only. Neither does $\langle \{N_{\langle \text{F} \rangle} \} \{V_{\langle \text{III} \rangle} \} \rangle$ describe a cumulative feature – which would be notated $\langle \{N_{\langle \text{F} \rangle} \{V_{\langle \text{III} \rangle} \} \} \rangle$. $N_{\langle \text{F} \rangle}$ takes over the VDS, turning it into an NDS in which $V_{\langle \text{III} \rangle}$ is present, but unordered vis-à-vis $N_{\langle \text{F} \rangle}$ and therefore unexpressible, whereas $N_{\langle \text{F} \rangle}$ is realized as /A/ (in this location). We see here yet another way that there may be a special paradigmatic relation between the VN and the finite verb: the VDS of the latter may serve as a hosting site for N in the former.

Unlike the VDS, however, the NDS is ordered after the *second* root consonant. Prefixing *mu-* – semantically a rather empty prefix, it seems – and simultaneously annexing /m/ to the root as a C_1 can now be understood as a device to comply with this property. (Such devices are used elsewhere in the grammar, e.g. adding or pruning one or two consonants in the broken plurals of nouns having too short – less than 3 C – or overlong – more than 4 C – roots. See Kihm 2003.) The CLR of *mukaataba* thus ends up being as in (35), which takes $N_{\langle \text{F} \rangle}$ splitting into account :

(35) $\{w \langle_{\text{r}} m, k, \langle \{N_{\langle \text{F} \rangle} \} \{V_{\langle \text{III} \rangle} \} \rangle, t, b, \langle \{N_{\langle \text{F} \rangle} \} \} \} \}$ CA *mukaataba* ‘act of writing to s.o.’
(Form III *kaataba*)

5.2.5 Form VI VNs

A close associate of Form III is Form VI, e.g. *taqaataluu* ‘they fought each other’, which seems to be derived by prefixing *ta-* to Form III.²⁹ The related **taCaaCuC** VNs (e.g., *taqaatul* ‘mutual fight’) include an initial syllable identical to the prefix of the finite verbs. They would therefore seem to be formed like Form II and IV related VNs (see *taksiir* and *’ijlaas* above) rather than like *mukaataba*. Yet, the surface long vowel realizing the NDS occurs following the second consonant from the beginning of the word, which we can only explain by assuming that, contrary to what is the case in *taksiir* and *’ijlaas*, but like *mu-* in *mukaataba*, the /ta/ syllable is annexed to the root, whereas $V_{\langle \text{VI} \rangle}$ is unordered, as shown in (36):

²⁹ If so, it means that *ta-* realizes $V_{\langle \text{VI} \rangle}$, while the surface long vowel realizes $V_{\langle \text{III} \rangle}$, a double exponence that does not contravene any principle of the present framework. As this is not a study of derived verbs, however, but of their related VNs, I will refrain from going any further into this matter.

(36) $\{w \langle_{\text{r}} t, q, \langle\{N\}\rangle, t, l \rangle \{V_{\langle VI \rangle}\}\}$ CA *taqaatul* ‘mutual fight’ (Form VI *taqaataluu*)

Such a possible annexation by the root confirms the non-significance of intrusive initial syllables in those VN templates that show them, even though these syllables correspond to meaningful prefixes in finite derived Forms, directly related or not – not always, though, as shown by *mu-* in Form III related *mukaataba*.

5.2.6. Form VII VNs

Moving on to Form VII, we find that *inkasara* ‘it broke’, for instance, is formed quite in the same way as Form IV ‘*ajlasa* ‘he seated’: $V_{\langle VII \rangle}$ fills up the LFS and is realized as an *n-* prefix that requires an epenthetic /i/ in front of it; the bare root follows:

(37) $\{w \langle\{V_{\langle VII \rangle}\}, \langle_{\text{r}} k, s, r \rangle\rangle\}$ CA Form VII *inkasara* ‘it broke’

We therefore expect the VN to be formed similarly as well, with the prefix present as an initial syllable and not annexed to the root. This is not exactly what we find, however. The VN related to (37) is *inkisaar* ‘action of breaking’. If the LFS of this form was filled up by a CV syllable as in *taksiir* and ‘*ijlaas*, we would expect /n/ to dock to the onset, yielding the wrong form **nikisaar*. It seems therefore we have to assume that in this case the prefix is shared as such by the VN and the finite verb, as shown in (38), to be compared with ‘*ijlaas* as represented in (28):

(38) $\{w \langle\{V_{\langle VII \rangle}\}, \langle_{\text{r}} k, s, \langle\{N\}\rangle, r \rangle\rangle\}$ CA *inkisaar* ‘action of breaking’ (Form VII *inkasara*)

Perhaps this peculiarity of Form VII is related to Guerssel & Lowenstamm’s (1990) conjecture that *n-* is actually the only really meaningful prefix in CA.

5.2.7. Form VIII VNs

Form VIII, e.g. *ijtama^cuu* ‘they convened’, is special in that it involves *ta* as an infix between C_1 and C_2 . According to Guerssel & Lowenstamm (1990), *-ta-* realizes the VDS, and there is no need for metathesis as in McCarthy’s (1981) account. The CLR of *ijtama^cuu* is as in (39):

(39) $\{w \langle_{\text{r}} j, \langle\{V_{\langle VIII \rangle}\}\rangle, m, ^c, \langle\{N \{NUM\}\}\rangle\rangle\}$ CA Form VIII *ijtama^cuu* ‘they convened’

The related VN is *ijtimaac* ‘meeting’, in which the consonant of the infix shows up. The problem with this form is that surface /aa/ realizing the NDS occurs after the root C_2 /m/, which suggests that the remnant of the infix, /t/, is not integrated to the root. This would make it the exponent of $V_{\langle VIII \rangle}$ in the VN, thus violating the templatic constraint against having more than one internal site in the root site. This constraint seems well grounded, however, so allowing it to be infringed would come at a high cost to the theory.

To avoid this I will lean on previous results and assume that /t/ as well is a mere legacy of finite Form VIII, devoid of meaning and transparent as far as root consonants count is concerned. Because of this, the NDS does follow C_2 in *ijtimaac*, and $V_{\langle VIII \rangle}$, not expressed by /t/, is unordered.

The lone consonant /t/ thus behaves like /’i/ and /ta/ in already examined VNs. There is a difference, though: whereas the former are ordered before the root, and simply passed over, /t/ is interposed between root consonants. This is why I say it is transparent.³⁰ Lest this move be seen as exceedingly *ad hoc*, I refer to Kihm (2004) where it is shown that consonant

³⁰ There is a phonological analogy here : some segments are transparent to, say, vowel harmony, others are opaque and block vowel harmony.

transparency is needed to account for some broken plural forms as well.³¹ We shall find more uses for it in the following. In (40) below, /t/ is italicized to indicate transparency:

(40) {_W ⟨_ʔ j, t, m, ⟨{N}⟩, °⟩ {V<VIII>}} CA *ijtimaa*^c ‘meeting’ (Form VIII *ijtima^cuu*)

5.2.8. Form IX VNs

Form IX is also special. Not only is it limited to so-called “adjectives of colours and defects” – salient qualities – such as ‘*aswad*’ ‘black’ or ‘*aTraš*’ ‘deaf’, but its distinctive feature is reduplication of the *third* root consonant, as shown in *iswadda* ‘he was/became black’ and *iTrašša* ‘he was/became deaf’. Initial /i/ is epenthetic, as usual. Carrying over to Form IX our account of Form II (see [10]) we are led to hypothesize an exceptional location for the VDS, as it encompasses C₃ rather than C₂:

(41) {_W ⟨_ʔ s, w, ⟨{V<IX> d}⟩⟩ CA Form IX *iswadda* ‘he was/became black’

VNs related to Form IX admit of one template, **CCiCaaC**, in which C₃ and C₄ are identical (cf. *iswidaad*, *iTriššaš*). This suggests the VN is actually formed on the 4C root produced by derivation, namely ⟨_ʔ s, w, d, d⟩ for *iswidaad*. But again we have to assume exceptional location following C₃ for the NDS realized as /A/ in C, as shown in (42):

(42) {_W ⟨_ʔ s, w, d, ⟨{N}⟩, d⟩ {V<IX>}} CA *iswidaad* ‘fact of being/becoming black’ (Form IX *iswadda*)

Here, what sets the connection between the VN and the finite verb is the exceptional location of the derivational site shared by both forms. Needless to say, I have no account for the exceptionality itself.

5.2.9. Form X VNs

Form X, e.g. *istaḥsana* ‘he considered preferable’, seems to raise no problem, since it is entitled to the same analysis as Forms IV and VII, namely (43):

(43) {_W ⟨{V<X>}, ⟨_ʔ ḥ, s, n⟩⟩ CA Form X *istaḥsana* ‘he considered preferable’

V<X> filling the LFS is realized as *sta-*. (See Guerssel & Lowenstamm 1990 for a finer analysis of this prefix.) The VN template **stiCCAC** (e.g., *istiḥsaan* ‘preference’) looks similar to the **ʔiCCAC** and **nCaCAC** templates of Forms IV and VII as well. There is a small hitch, however, namely the clearly epenthetic vowel /i/ following /st/ in the VN instead of /a/ in the finite verb. Given our account of Form VIII VNs (see [40]), this discrepancy alerts us to the possibility that /st/ in *istiḥsaan* may be the same sort of transparent phonological carry-over as /t/ is in *ijtimaa*^c. If that is so, the proper representation of *istiḥsaan* is as in (44):

(44) {_W ⟨_ʔ s, t, ḥ, s, ⟨{N}⟩, n⟩ {V<X>}} CA *istiḥsaan* ‘preference’ (Form X *istaḥsana*)

5.2.10 Form V VNs

We finally come to Form V, as in *tasallama* ‘he received’ and *takallama* ‘he spoke’, and its two VN templates : **taCaCCuC** as in *tasallum* and **tiCiCCAC** as in *tikillaam*. Form V is similar to Form VI insofar as it seems to consist in Form II (instead of Form III) to which *ta-* is prefixed. I will therefore assign it the same CLR as to Form VI (see fn. 29), namely (45) in which V<V> in the LFS is realized as *ta-* whereas internal V<II> geminates C2:

³¹ At any rate, I am confident it is less *ad hoc* than metathesis.

(45) $\{w \langle \{V_{<V>}\}, \langle_{\text{ʔ}} k, \langle \{V_{<II> l}\}, m \rangle \rangle\}$ CA Form V *takallama* ‘he spoke’

The distinctive feature of the VN *tikillaam* is the fact that initial /t/ is followed by what clearly is an epenthetic /i/. This, as in previous cases, suggests that initial /t/ in the VN is not the prefix that appears in the related finite verb, but a phonological carry over. (Recall it was precisely the absence of /i/ after /n/ in Form VII related *inkisaar* that led us to conclude that /n/ here is the prefix.) Likewise, the location of the surface /aa/ realizing the NDS indicates that the double consonant is also a legacy from Form II in which the ‘double’ is transparent. All this concurs to have $V_{<V>}$ unordered in the VN, hence the following CLR, in which initial CV is realized as /ti/:

(46) $\{w \langle C, v, \langle_{\text{ʔ}} k, l, l, \langle \{N\}\rangle, m \rangle \rangle \{V_{<V>}\}$ CA *tikillaam* ‘action of speaking’ (Form V *takallama*)

The other template, **taCaCCuC**, presents us with a choice given the timbre of the vowel following initial /t/: either /ta/ is the prefix *ta-* of the related finite verb, or /a/ is also epenthetic, i.e. provided by phonology, but it is low for some reason, perhaps harmony with the /a/ following C_1 . Since I cannot think of any good reason for proposing diverging accounts for the two templates I adopt the second alternative. The CLR of *tasallum* is therefore the same as that of *tikillaam* above. The real difference is in realization: *tasallum* is like **CuCuC** broken plurals (cf. *kutub* ‘books’), i.e. with the glide /U/ filling the nucleus rather than the onset of the extra syllable provided by NDS activation.

5.2.11. VNs of 4C verbs

Let us now look at the five VN templates related to quadrilateral (4C) verbs. No root VNs are to be found among them, which does not come as a surprise. Interestingly, no comparison with 4C broken plurals seems to be feasible. One template, however, namely **CiCCAC** associated with Form I₄, as in *dihraaj* ‘action of rolling’ (cf. *dahraja* ‘he rolled’) can be regarded as a broken plural form (**CiCAC**) provided we assume that either one of C_2 or C_3 is transparent with respect to NDS location. For this we may lean upon the grammarian’s old insight that 4C verbs in Form I are never basic, even when not obviously reduplicative like *tamtama* ‘he stammered’, but they include an intrusive consonant, often represented by /l/ or /r/. Given this, we may assign *dahraja* the following CLR:

(47) $\{w \langle_{\text{ʔ}} d, h, \langle \{V_{<I4>}\}, j \rangle \rangle\}$ CA Form I₄ *dahraja* ‘he rolled’

In (47), $V_{<I4>}$ hosted by the VDS is realized as /r/, the intrusive consonant, actually a (probably fossilized) derivational morph. It is this /r/ that is passed over to the related VN, in which it is transparent, so the NDS effectively follows C_2 /h/, as shown in (48):

(48) $\{w \langle_{\text{ʔ}} d, h, r, \langle \{N\}\rangle, j \rangle \{V_{<I4>}\}$ CA *dihraaj* ‘action of rolling’ (Form I₄ *dahraja*)

Except for $n_{<F>}$ split and the different vowelings, the same account holds of the alternate template manifested by *dahraja(tun)*.

For derived Forms of 4C verbs, it seems we must assume two verbal DSs, one for the intrusive consonant, perhaps to be viewed as a fossilized derivation as we have just seen, and one for the active derivation. The intransitive counterpart of *dahraja*, *tadahraja* ‘it rolled’, therefore has the following CLR, in which $V_{<II4>}$, the active derivation, fills up the LFS and is realized as *ta-*, whereas the internal fossilized derivation yielding /r/ is notated as $V_{<4>}$:

(49) $\{w \langle \{V_{<II4>}\}, \langle_{\text{r}} d, \text{h}, \langle \{V_{<4>}\}, j \rangle \}$ CA Form II₄ *tadaħraja* ‘it rolled’

The VN *tadaħruj* ‘rolling’ is then formed partly like the Form V VN *tasallum* (for initial /ta/ and the NDS realized as short /u/) and partly like *dihraaj* (for /r/ transparency):

(50) $\{w \langle c, v, \langle_{\text{r}} d, \text{h}, r, \langle \{N\}\}, j \rangle \{V_{<II4>}\}$ CA *tadaħruj* ‘action of rolling’ (Form II₄ *tadaħraja*)

Form III₄, e.g. *ixranTama* ‘he raised the nose, he was proud’, is characterized by a nasal infix, perhaps the same morph as appears as a prefix in Form VII (cf. *inkasara* ‘it broke’) in addition to intrusive /r/. The root here is $\langle_{\text{r}} x, T, m \rangle$, manifested by *xaTm* ‘nose’. As we wish to abide by the templatic constraint (22), I would suggest that both intrusive /r/ and the infix realize the internal VDS hosting $V_{<III4>}$, as shown in (51):

(51) $\{w \langle_{\text{r}} x, \langle \{V_{<III4>}\}, T, m \rangle \}$ CA Form III₄ *ixranTama* ‘he raised the nose, he was proud’

It follows from this that both /r/ and /n/ are transparent in the related VN *ixrinTaam*, hence the location of the NDS after the actual root $C_2 /T/$:

(52) $\{w \langle_{\text{r}} x, r, n, T, \langle \{N\}\}, m \rangle \{V_{<III4>}\}$ CA *ixrinTaam* ‘action of raising the nose, pride’ (Form III₄ *ixranTama*)

Finally, Form IV₄, e.g. *iqša^carra* ‘he shuddered’, seems to involve gemination of the intrusive consonant, since the basic root appears to be $\langle_{\text{r}} q, \text{š}, \text{c} \rangle$ manifested by *qaša^ca* ‘he scattered’. As there is no obvious relationship between ‘to scatter’ and ‘to shudder’, I will simply assume that /r/ in this form, although perhaps intrusive and derivational originally, is a root consonant like the other consonants, so that Form IV₄ results from the VDS exceptionally encompassing C_4 , as shown in (53):

(53) $\{w \langle_{\text{r}} q, \text{š}, \text{c}, \langle \{V_{<IV4> r}\}\rangle \}$ CA Form IV₄ *iqša^carra* ‘he shuddered’

Form IV₄ is thus similar to Form IX. So is the related VN *iqši^craar* (compare *iswidaad* in [42]):

(54) $\{w \langle_{\text{r}} q, \text{š}, \text{c}, r, \langle \{N\}\}, r \rangle \{V_{<IV4>}\}$ CA *iqši^craar* ‘shudder’ (Form IV₄ *iqša^carra*)

There is an alternative form *quša^criira* ‘shudder, tremor’ which, if it is indeed a VN, seems to involve $N_{<F>}$ split and realization of the NDS as /i/ in the C slot.

6. Summary

Two conclusions result from the foregoing review. First, VNs are either phonologized bare roots (root VNs) or they are formed like existing or possible broken plurals. Despite ancient ideas of the matter (see, e.g., Barth 1904), this should not be taken as implying any inherent (semantic) relation between the two formations in the sense that, for instance, broken plurals interpreted generically or massively and VNs would both refer to some sort of ‘abstraction’ of

the individual meaning of the singular noun or finite verb.³² I am not saying that the idea does not make sense, only that it is doomed to remain a conjecture, which takes away most of its potential interest. In contrast, the formal similarity of VNs and broken plurals does point to a significant fact of CA morphology, namely the lack of derivational suffixes such as English *-hood* and, given the strong restrictions on external ('sound') pluralization, the overall prevalence of internal morphological processes.³³ VNs and broken plurals resemble each other because they exploit the same device, and they do so because there really isn't any other device available – prefixation being also a limited resource given the paucity of prefixes and the vagueness of their meanings, to say the least.

The second conclusion is about the connections between VNs and the corresponding finite verbs. Here, two levels must be distinguished: template formation and template selection. As far as the latter is concerned, we are dealing with the association of certain templates with certain aspects of the verb's meaning. Again, Form I (so-called 'basic') must be set apart from derived Forms.

Two aspects of the meaning of Form I seem to play a role. One is the verb's inherent, 'lexical' meaning, hence the VN templates apparently specialized for verbs denoting refusal, or sickness, or violent motion, etc. Here, more research is necessary to ascertain whether such connections are real, or are artefacts due to partial observation, as I am inclined to suspect, perhaps wrongly.

The other connection, although certainly not 100%, seems much more robust, and it is with the aspectual class of the verb as evidenced by the timbre of the second vowel in the Perfective.³⁴ Reviewing and interpreting the facts listed above, we see the following: **CaCC** VNs are associated with transitive verbs having agentive subjects and /a/ in v_2 (cf. *qatl* 'murder' / *qatala* 'he killed') and with transitive verbs whose subjects are non-agentive and v_2 is /i/ (cf. *fahm* 'understanding' / *fahima* 'he understood'); **CuCuuC** VNs correspond to unergative verbs with /a/ in v_2 (cf. *juluus* 'seating' / *jalasa* 'he sat'); **CaCaC** VNs go with unaccusative verbs with /i/ in v_2 (cf. *farah* 'joy' / *fariha* 'he rejoiced'); **CaCaaCa(tun)** and **CuCuuCaa(tun)** associate with stative verbs (cf. *saraawa(tun)* 'generosity' / *saruwa* 'he is generous', *suhuula(tun)* 'smoothness' / *sahula* 'it is smooth').

Such a connection cannot be observed with derived verbs, since aspectual classes are not distinguished in them. Neither do we observe a correlation with the inherent meaning. In fact, VN templates are 'simply' paired with Forms, keeping the meaning specification (if any) the derivation adds to the simple verb (causativeness, intensification, etc.). There may be an explanation for this difference between Form I and derived Forms VNs. It resides in the other connecting level we have to consider, namely template formation to which we now turn.

³² Barth's reasoning may be reconstructed thus: in the same way that the abstract noun 'manhood' (CA *rujuula*) is akin to the generic plural 'men' (CA *rijaal*) meaning anything that has the property of being a man, the VN '(action of) killing, murder' (CA *qatl*) has the potentiality to refer to any action that has or had or will have the property of being an instance of killing of anyone by anyone. Given the historical bias of linguistics in his time, Barth therefore conjectures that VNs were actual plurals of verbs in the prehistory of Semitic languages. (Not the other way around because plurality is taken to be more primitive than abstraction.) Moreover, broken plurals were not considered to be true plurals, a view that is clearly expressed in Wright (1, 233): "As regards their meanings, the plurales fracti differ entirely from the sound plurals; for the latter denote several *distinct* individuals of a genus, the former a number of individuals viewed *collectively*, the idea of individuality being wholly suppressed." (Emphasis in the original.) This view stumbles against the fact that broken plurals are and seem always to have been used to express true plurality, if only because the overwhelming majority of the nouns and adjectives that form them do not form sound plurals in addition, so there is no possible contrast.

³³ This is no longer entirely true of Modern Hebrew, as it makes productive use of derivational suffixes, especially *-ut* (e.g., *balšanut* 'linguistics' < *balšan* 'linguist, philologist' < *baleš* 'to inquire'). Yet, their form reveals these suffixes for former feminine endings, which links such derivations to, e.g., *ʔinsaaniyya(t)* 'humanity', the feminine of *'insaaniyy* 'human', from *'insaan* 'human being'.

³⁴ I take aspectual in a very broad, probably partly inadequate sense.

Viewed from the widest perspective, VN formation consists in combining the root with the two categorial features N and V. Hence, VNs do not pertain to the verbal inflectional paradigm which includes only V.³⁵ What they do belong to is the derivational paradigm. Such membership does not imply that VNs are derived from a finite verb form. In this respect, the Basris were right against the Kufis.

VNs related to Form I belong to a derivational paradigm the entry point of which is the bare root. Some, the root VNs, realize this bare root plus one or two vowels supplied by phonology (in the sense that they are not represented at any other level than the phonological description). Neither N nor V have dedicated exponents. Others, the broken VNs, result from the activation of the nominal derivational site (NDS) located between root C₂ and C₃ and hosting the categorial feature N. In phonology the NDS is realized as an extra syllable, and N as a glide (/A/, /U/, or /I/) docking to the onset or the nucleus of the said syllable. (V has no dedicated exponent.) Form I VNs therefore consist in the root with or without the addition of an exponent of N.

VNs related to derived Forms II to X, on the other hand, may retain elements of these Forms. They therefore are members of a derivational paradigm the entry point of which is the root possibly augmented with affixes realizing the verbal derivation they are paired with – or another derivation, as is the case with Form II VNs which ‘take’ their prefix from Form V. As we saw, however, these affixes generally have no morphological value in the VNs. Actually, they are no longer verbal affixes, but meaningless and possibly transparent elements extending the basic root. Only N in the NDS is given an exponent, whereas V_{<x>} (x the Form’s number) is unordered and unrealized. (Form VII VNs seem to be the only exception, as it retains *n-* as a prefix realizing V_{<VII>}.)

The fact that VNs of derived Forms may retain elements of these Forms in the way they do demonstrates two things. On the one hand, it confirms that VNs are not derived from verb forms because, if they were, we would expect the carry-overs to be more regular and systematic than they are. That is, we are not dealing with word-to-word derivation, but VNs are built from a more abstract object, namely the (possibly extended) root. On the other hand, however, such inheritance points to the existence of what I have called paradigmatic side relations, meaning that although not part of the verbal paradigm, VNs are related to it both formally and semantically.

As already mentioned, Form I and the derived Forms stand in sharp contrast here. Only with the latter is a significant formal connection apparent, and it is due to the sharing of elements added to the root by the VN and the related finite verb (cf. *ijtimaad^c* vs. *ijtama^cuu*, *iswidaad* vs. *iswadda*, etc. – shared elements beyond the root are set in bold). With Form I, in contrast, the formal connection is trivial since only the root can (and must) be common to the VN and the finite verb. The semantic connection stems from the fact that the categorial feature V is present on both sides with the same value. In Form I this value is ‘aspectual’ and there is some evidence, although not quite conclusive, that the root’s inherent meaning is accessible as well. It follows that there may indeed be a correlation between a given VN template and a given aspectual or semantic verb class, even though it is far from systematic, as we saw. In derived Forms, in contrast, V’s value only refers to the semantic contribution of the derivation itself, superseding the aspectual value of Form I as well as the root’s inherent meaning, supposing it to be ever relevant.

³⁵ It may include N as an agreement feature, but that is another matter.

7. Consequences for the meaning and syntactic properties of VNs

7.1. The Voice indetermination of VNs

Beyond such form-meaning correlations VNs still raise two issues of a semantic and syntactic nature. The first, already mentioned, is that VNs are not specified for Voice, contrary to the verbs they are related to. It is well known, for instance, that a VN-headed ‘construct state’ noun phrase such as (55) out of context may be interpreted either actively as ‘Zayd’s kissing (s.o.)’ or passively as ‘Zayd being kissed (by s.o.)’.³⁶

- (55) taqbiilu zaydin
 kissing-NOM.DEF Zayd-GEN
 Zayd’s kissing

No such ambiguity exists in finite verbs, as shown by (56) and (57):

- (56) qabbala zaydun ’abaahu
 kiss.3SG.MASC.PF Zayd-NOM father-ACC-POSS.3SG.MASC
 Zayd kissed his father.
 (57) qubbila zaydun
 kiss.PASS.3SG.MASC.PF Zayd-NOM
 Zayd was kissed.

Participles have distinct active and passive forms as well (cf. *muqabbilun* ‘kissing’, *muqabbalun* ‘kissed’).

Similar facts of course occur in other languages. I am referring in particular to the so-called ‘subjective’ and ‘objective’ Genitives in expressions such as Latin *metus hostium* and its English translation ‘the fear of the enemies’, or *opiniones deorum* ‘the gods’ thoughts’ or ‘the thoughts about the gods’. This also may be considered a case of Voice indetermination of the deverbal nouns *metus*, *fear* (compare *The enemies fear* vs. *The enemies are feared*) and *opinio*. However, the verbal status of such nouns has been questioned, so it is not clear why they should have anything to do with Voice distinctions in the first place.³⁷ Arabic VNs, in contrast, are unquestionably verbal in their syntactic behaviour, as we shall see, and there is a principled reason that they cannot be specified for Voice.

Passivization in Arabic is an internal operation like most morphology in the language. Outwardly, it consists in realizing the vowel melody of the base as /u... i/ in the Perfective (cf. *lubisa* ‘he was dressed’, *qubbila* ‘he was kissed’, *duħrija* ‘it was rolled’), as /u... a/ in the Imperfective (cf. *yulbasu* ‘he is dressed’, *yuqabbalu* ‘he is kissed’, *yudaħraju* ‘it is rolled’), irrespective of the thematic vowel of the active in Form I (see [11]).³⁸ It may not be an accident that /u/, considered the distinctive vowel of the Passive, is also the thematic vowel of stative verbs (cf. *kabura* ‘he is big’).

In the present framework, this means that the Passive vowel melody realizes an internal site identified by [Voice] that includes the tense-aspect feature [T] (if present, that is non-default, that is IPF), and which encompasses all vowels slots. Since nonpassive verbs are not specifically marked as such, I assume the Voice site is not activated in them. In other

³⁶ I transcribe case endings and the tanwīn from now on. Notice that person’s names like *zaydun* take the tanwīn although they are inherently definite.

³⁷ In the sense that *fear* and *opinio* look more like result nominals than like complex event nominals to use Grimshaw’s (1990) terminology.

³⁸ The base consists in the consonantal root plus the interposed vowels, called the vowel melody.

words, [Voice] is always interpreted as Passive. Active verbs are simply voiceless. The CLR of, e.g., *qubbila* ‘he was kissed’ is thus the following:

(58) $\{W \langle \mathcal{R}/B \ q, \langle \{Voice\} \langle \{V_{<II>} \ b \} \rangle, l \rangle \}$ CA Form II Passive *qubbila* ‘he was kissed’

The Voice site opens following C_1 /q/ and it closes following C_2 /b/. It is notated below the line to indicate that it encompasses not the root consonants, but the relevant slots of the interposed vowel melody, i.e. v_1 and v_2 realized /u/ and /i/ because of [Voice]. (Final /a/ does not seem to have morphological value.) This is why I label the maximal ordered set $\langle \mathcal{R}/B \ \rangle$, $B =$ base, because it contains sites acting on the root consonants and on the vowel melody. [T] is not present in (58) since its value is default. Likewise, Person, Number and Gender of the subject are all default, therefore not notated.

Consider now the Imperfective *yuqabbalu* ‘he is kissed’. The positional instructions for the Voice site now are that it opens following the first consonant of the *inflected* form, namely /y/ in the given example, and that it closes before the last consonant of the root (see *yulbasu*, *yudaħraju* – again final /u/ is probably meaningless). That the Voice site should align on the inflected form is only to be expected, since it includes non-default [T], and the initial prefix ought probably to be taken as realizing non-default, imperfective [T], possibly with additional non-default values for Person, Number and Gender of the subject. Hence the following CLR for *yuqabbalu* ‘he is kissed’:

(59) $\{W \langle \langle \{Voice \ T\} \langle \mathcal{R}/B \ q, \langle \{V_{<II>} \ b \} \rangle, l \rangle \rangle \}$ CA Form II Passive *yuqabbalu* ‘he is kissed’

The [Voice-T] site opens where the RFS does. Voice is realized as the vowel sequence /u, a⁺/, where /u/ is ordered after the first consonant of the word (W) and ‘+’ is the Kleene plus, so /a⁺/ means one or more occurrences of /a/ following successive consonants.³⁹ Non-default [T] with default values for so-called ‘phi’ features is realized as y-. If [Voice] was not present, [T] would also operate on the vowel melody (cf. *yuqabbilu* ‘he kisses’), but [Voice] apparently superimposes its own melody /u, a⁺/ to the Imperfective melody /u, a, i/.

The realization of (59) proceeds as in (60):

(60) $((((\langle T \dots \rangle f_c \rightarrow /y, /) \times (\langle q, \langle \rangle, l \rangle f_c \rightarrow /k, \langle \rangle, b, /) \times (\langle \{V_{<II>} \ b \} \rangle f_{cv} \rightarrow \text{Geminate})) f_c \rightarrow /q, b, b, l, /) \times (\langle \{Voice\} \dots \rangle f_v \rightarrow /u, a^+ /)) f_{cv} \rightarrow /yuqabbalu /)$ CA Form II Passive *yuqabbalu* ‘he is kissed’

It consists in five steps: (i) the partial realization function f_c applies to [T], returning y-, given phi-feature defaultness; (ii) f_c reapplies to the members of the root site that are not also members of another site, returning C_1 /q/ and C_3 /l/; (iii) f_c applies to the VDS and returns geminated /b.b./; (iv) the partial realization function f_v applies to [Voice], returning the vowel melody; (v) the complete realization function f_{cv} applies to the composition of all previous realizations, returning the nearly fully specified MTS /yuqabbalu/.

Coming back to VNs, we know they have the form of bare roots or of broken plurals. In the first case, morphology does not concern itself with vowing at all. In the second case, only the long or short vowel that manifests the NDS (depending on the linking of the spell-out glide to the consonant or the vowel slot) is morphologically relevant. As a consequence, the

³⁹ We can let **yuqababalu* be generated by this procedure, since it will automatically be corrected to *yuqabbalu* in phonology.

vowel melody as such cannot be an exponent in VNs. Therefore VNs must be unspecified for Voice, the exponent of which is the vowel melody.

7.2. VNs and their arguments

Being verbal, i.e. including [V], VNs must have subjects and can have objects. Argument grids are associated with them, in other words.⁴⁰ This was already implicit in the previous discussion about Voice, since another way of saying that VNs are not specified for it is to observe that *zaydin* in *taqbiilu zaydin* can be interpreted as the subject in the active reading, the object in the passive reading, of *taqbiilun* ‘action of kissing’. When both a subject and an object are present together, however, no such double reading is of course possible, and VNs are unambiguously active. The usual state of affairs, then, is for the subject to be marked with the Genitive case and the object with the Accusative case, as in the following example adapted from Haywood & Nahman (1965: 330):

- (61) ^oajibtu min taqbiili zaydin ’abaahu
 be.astounded.PF-1SG from kissing-GEN Zayd-GEN father-ACC-POSS.3SG.MASC
 I was astounded at Zayd kissing his father.

What this case marking pattern suggests is that the VDS and the NDS included in *taqbiilun* play a syntactic role parallel with their morphological activity. I am not interested here in the precise syntactic structure of the noun phrase [taqbiil(i) zaydin ’abaahu]. (The Genitive of *taqbiili* is assigned to it by the preposition *min*.) It is enough for my purpose that, given the present analysis of VNs and the case marking pattern, it comes as a natural conclusion that *zaydin* in (61) is actually the argument of VN-internal [N], whereas *’abaahu* is the argument of VN-internal [V_{<II>}]. Verbal items assign Accusative to their objects. By ‘verbal items’ I mean not just ordinary finite verbs, but also various sorts of copulas (e.g. *kaana* ‘he was’, *baqiya* ‘he remained’, *Saara* ‘he became’, etc.), discourse particles such as *’inna* ‘truly’, *laa* meaning ‘there is no...’, and so forth. Since all these items have verbal (or predicative) force, it is only normal that the abstract verbal feature [V] should be endowed with the same case marking property as they are. Put differently, Accusative (*naSb*) in CA signals that an item including [N] in its CLR is the argument of another item including [V]. This argument is unambiguously interpreted as the object, since subjects are not arguments of verbal items, but of the VPs (or predicates) the verbal items head.

The Genitive (*jarr*), on the other hand, signals that an item including [N] is the argument of another item also including [N] (or including neither [N] nor [V], if that is the case with prepositions). There are three possibilities. If the item, a primitive noun, includes only [N], the argument is interpreted as standing in some relation to it, usually possession (e.g., *saa’atu l-mar’ati* ‘the woman’s watch’), also part to whole (e.g., *qiT’atu lahmin* ‘a piece of meat’), and others.

If the item, a VN, includes [N] and [V], but there is just one argument, this argument is [N]’s argument, as shown by the Genitive marking, but its interpretation is ambiguous between subject and object as we saw with *taqbiilu zaydin* ‘Zayd’s kissing’. In itself the ambiguity is unsurprising since the genitive relation is largely underspecified across the board: cf. *Suuratu zaydin* ‘Zayd’s picture’, meaning either the picture showing Zayd or the picture Zayd made – or the picture Zayd owns, and so forth. The question we must ask is why

⁴⁰ In HPSG terms, VNs like Vs are [SPR –], meaning they need a specifier (see Sag, Wasow & Bender 2003, Chapter 3). Given the V/NSO character of the language, all specifiers in CA, of verbs as well as of nouns, are ordered to the right of the head by a linear precedence (LP) rule. Care should be taken to distinguish [V] and [N], the abstract features, from the syntactic categories V and N. What follows is based mainly on Wright’s exposition of the facts. Also see Badawi, Carter & Gully (2004: 237ff.)

cannot [V] govern the unique argument. Why is **taqbiilu zaydan*, with *zayd* in the Accusative, meaning ‘the action of kissing Zayd’, ungrammatical? The answer, I think, is that [N] and [V] do not stand on an equal footing in VNs: [N] is the head insofar as it imparts its category to the VN, a *noun* with verbal properties, not the other way around. Therefore, [N] overrides [V] if there is only one argument to govern and to case-mark.

Only if there are two arguments is [V] allowed to play a syntactic role. It then takes as its own argument the noun phrase denoting the entity affected by the event, while [N] selects the agent or the experiencer as in (62) where the possessive pronoun suffixed to *maḥabba* ‘love’ (a VN form I did not review) functions as the Genitive argument of [N] in *maḥabba*, hence as the subject (specifier) of the predicate [maḥabbatu l-baladi] ‘(the) love of the country’.⁴¹

- (62) maḥabbatii l- balada
 love.NOM-POSS.1SG D-country.ACC
 my love for the country

The fact that a pronoun suffixed to the VN must be interpreted as the subject (in this pattern, but see below) has a consequence when the object is pronominal. This pronoun cannot attach to the VN, which would end up having two subjects. A so-called ‘detached object pronoun’ suffixed to the particle *’iyyaa*, which takes charge of semantic role assignment and case marking, is then required, as in (63):

- (63) ^cajibtu min taqbiili zaydin ’iyyaahu
 be.astounded.PF-1SG from kissing-GEN Zayd-GEN PTCL-3SG.MASC
 I was astounded at Zayd kissing him.

Suffixing the object pronoun to the VN would result in ungrammaticality (cf. **^cajibtu min taqbiilihi zaydin*). This is also the usual solution when both the subject and the object are pronominal as in (64):

- (64) ^cajibtu min taqbiilihi ’iyyaahu⁴²
 be.astounded.PF-1SG from kissing-GEN-3SG.MASC PTCL-3SG.MASC
 I was astounded at him kissing him.

However, Wright (2, 194ff) mentions that both pronouns may be suffixed to the VN in the order subject < object, as in his example (65), where I insert hyphens for clarity:

- (65) ḥubb-ii-hi
 love-1SG-3SG.MASC
 my love for him

This is not a problem, though, because the object pronoun – here /-hu/ ‘him’ realized [hi] (see fn. 42) – is then attached not to the VN, but to the phrase including the VN and its subject. The object pronoun is then no different from a noun in object position, except for being a phonological clitic. The qualification ‘phonological’ is crucial, because if *ḥubbiihi* was syntactically derived in a way implying the Mirror Principle, we would expect the reverse

⁴¹ Recall that the unique argument must be marked Genitive, hence *baladi* here as against *balada* in (62). Although not overtly Genitive, suffixed *-ii* ‘my’ counts as a Genitive.

⁴² The suffixed pronoun *-hu* becomes *-hi* in *taqbiilihi* (compare *’iyyaahu*) because of assimilation to the preceding [i].

order **ḥubbḥuyii*, since the object stands closer to the verb or VN than does the subject, and (generative) syntax does not recognize a phrase consisting of the verb and its subject. This contributes to showing that morphology does not necessarily mirror syntax – probably because they are not or quite indirectly connected.

The Genitive subject / Accusative object pattern is not the only possible one, however. Another pattern, ‘not uncommon’ according to Wright (2, 194ff), is Nominative subject / Genitive object so that (61) can be turned into (66):

(66) ^cajibtu min taqbiili ’abiihi zaydun
 be.astounded.PF-1SG from kissing-GEN father-GEN-POSS.3SG.MASC Zayd-NOM
 I was astounded at Zayd kissing his father.

In (66) *taqbiili ’abiihi* ‘kissing of his father’ is a Construct State Nominal, which is why *ʔabiihi* must not be separated from *taqbiili*, and the Nominative subject *zaydun* follows. We are thus dealing with a VOS construction. This pattern is also found, frequently, when the object is a pronoun, as in ^c*ajibtu min taqbiilihi zaydun*. Shall we take it as counterevidence to the claim that the Genitive argument must be interpreted as a subject, as it is in (61) and (62)?

The crucial observation, I think, is that this pattern is acceptable only if the subject argument is in the Nominative (*raf*⁶) (compare *^c*ajibtu min taqbiilihi zaydin* already mentioned). Now so-called Nominative clearly is the case assigned to nominal lexical items that are not the argument of a verbal or a nominal lexical item. It is the “elsewhere case”, always inherent, whereas the Genitive is always structural, and the Accusative is one or the other. Take a sentence such as (67):

(67) zaydun mu^callimun qadiimun
 zayd teacher-NOM-TW old-NOM-TW⁴³
 Zayd is an old teacher.

All items in this sentence are in the Nominative, *mu^callimun qadiimun* is the predicate phrase, and the subject *zaydun* is not the argument of a lexical item, but of the predicate phrase.

Consequently, in *taqbiil(i) ʔabiihi zaydun*, *zaydun* is not the argument of *taqbiil(i)*, but of the phrase *taqbiil(i) ’abiihi*, whereas *’abiihi* is the sole argument of *taqbiil(i)* itself. We are thus brought back to *taqbiilu zaydin*, in which, as we saw, *zaydin* is case-marked by [N], and [V] is syntactically idle because overridden by [N]. In principle, then, *’abiihi* should be ambiguous since *taqbiilu ’abiihi* ‘his father’s kissing’ in isolation means either that his father was kissed or that he kissed some unmentioned person. We are not in isolation, however, as there is another lexical item, *zaydun*, that is unambiguously marked as being the subject of the predicate ‘kissing his father’. The active reading of *taqbiil(i) ’abiihi* is therefore the only possible one. Naturally, the same reasoning applies when the object is a pronoun.

To summarize: in the VN – S_{GEN} – O_{ACC} pattern, [N] governs and case-marks the argument interpreted as the subject, while [V] governs and case-marks the argument interpreted as the object. I therefore call it the [NV]-pattern. In the VN – O_{GEN} – S_{NOM} pattern, [N] governs and case-marks the sole argument of the VN interpreted as the object, while the remaining item is (and must be) inherently case-marked as Nominative, hence interpreted as the subject of the predication. [N]-pattern is thus an adequate label for it.

Finally, there is a third and rare pattern, namely VN – S_{NOM} – O_{ACC} or VN – O_{ACC} – S_{NOM} (both orders are grammatical), such that one may report about one’s astonishment at Zayd kissing his father by uttering (68) or (69):

⁴³ TW = tanwîn.

- (68) ^ʕajibtu min taqbiilin ʾabaahu zaydun
 be.astounded.PF-1SG from kissing-GEN father-ACC-POSS.3SG.MASC Zayd-NOM
 I was astounded at Zayd kissing his father.
- (69) ^ʕajibtu min taqbiilin zaydun ʾabaahu
 be.astounded.PF-1SG from kissing-GEN Zayd-NOM father-ACC-POSS.3SG.MASC
 I was astounded at Zayd kissing his father.

The VN then behaves, to all intents and purposes, like a finite verb. The existence of such a pattern is not surprising, given our framework. It is the reverse of the [N]-pattern: [V] instead of [N] governs and case-marks the argument of the VN. Note that the absence of a Genitive argument of [N] entails that the VN appears with *tanwîn*, even though the noun phrase it heads cannot be said to be indefinite.

CA thus exploits, with varying degrees of frequency, all possibilities of argument licensing, by [N], by [V], or by both. (The fourth possibility, to wit by neither, is merely logical.) It is a merit, I believe, of the present analysis of VNs that it allows one to bring this systematicity to light.

8. Conclusion: sites, apophony, and templates

8.1. Site morphemes vs. apophony

In the theory of VN and broken plural formation upheld here, apophony plays no role, as we saw. Asfour (2001) had to use it to explain ‘short’ broken plurals such as *kutub* ‘books’ (sg. *kitaab*) or *ġuraf* ‘rooms’ (sg. *ġurfa*), which I have shown to be accountable by the same process of glide insertion in the NDS as ‘iambic’ broken plurals like *kilaab* ‘dogs’ (sg. *kalb*) or *nufiūs* ‘souls’ (sg. *nafs*). Given this, the domain of apophonic processes in Arabic appears to be very small indeed, perhaps limited to vowel alternations between Perfective and Imperfective verb forms, as in *labis* ‘he dressed’ vs. *yalbas* ‘he dresses’, *katab* ‘he wrote’ vs. *yaktub* ‘he writes’, *Darab* ‘he hit’ vs. *yaDrib* ‘he hits’, and *kabur* ‘he got big’ vs. *yakbur* ‘he gets big’ (see Guerssel & Lowenstamm 1996).

This is in keeping with the overall scarcity of the use of apophony as a morphological, i.e. meaningful process in natural languages. What counts as an apophonic process? I propose the following, restrictive definition whereby such a process must (a) express a meaning which nothing else expresses in the same unit ; (b) consist entirely in a feature shift within a segment. Condition (a) excludes cases such as the vowel alternation of French (*je*) *mène* [mɛn] ‘I lead’ vs. (*nous*) *menons* [mənɔ̃] ‘we lead’, since the [ɛ] / [ə] contrast means nothing in itself. It also excludes German umlaut as in *Füsse* ‘feet’, where the primary exponent of plurality is the *-e* ending.

In that case, there is a plausible phonological trigger, but there need not be one. For instance, the Papuan language Hua has an apophonic process such that the final vowel of a verb stem is fronted when the verb form includes a 2nd or 3rd person dual or plural exponent as shown in (70) and (71) (see Haiman 1998: 547-548):

- (70) Do-ga- ne.
 eat-FUT-2SG
 You will eat.
- (71) De-ga- 'e.
 eat-FUT-2/3DUAL
 You/they two will eat.

The *o/e* alternation in the stem *do/e* is not obviously triggered by anything in the phonological environment. However, it is not the sole exponent of the contrast singular / dual in the given example, which is primarily expressed by the pronominal suffix. Therefore, Hua ablaut (to use Haiman's term) ought not to be considered morphemic in the sense I am proposing. Rather it is one of the numerous examples of phonologically conditioned apophony where the conditioning segment has been lost in the course of history (see Haiman 1998: 539).

Another case in point are Fula initial consonant mutations, for which the classical description in English is Arnott (1970). In pairs such as *gor-ko* /man-CL1/ 'man' vs. *wor-Be* /man_{MUT}-CL2/ 'men', the continuant grade mutation /g/ → /w/ is correlated with the presence of the plural class marker (CL2) *-Be*, but it is not the primary exponent of plurality. Insofar as it may be viewed as an apophonic process, it is thus not a morphemic apophonic process. Here too, as in Hua, the ultimate explanation is historical.

Condition (b), on the other hand, clearly rules out CA VN and broken plural formation as a possible apophonic phenomenon, for the simple reason that, if the present analysis is correct, such formation does not affect any segment of the finite verb or singular form of the noun. As put, however, the condition will also exclude cases such as Terena nasalization which Spencer (1998: 132-133) mentions as an apophonic process – cf. *emo'u* 'his word' vs. *emō'u* 'my word', *piho* 'he went' vs. *^mbiho* 'I went'. True, nasality appears as the sole exponent of the (complex) feature [1SG]. Yet, (pre)nasalization cannot be considered a feature shift within a segment; it is the result of a nasal autosegment /N/ – the exponent of [1SG] – linking to a plosive (/p/ → ^mb/) or, in the absence of one, to a vowel (/o/ → /ō/). In other words, /N/ is a morpheme the only peculiarity of which is not enjoying its own position in the phonological sequence. Terena nasalization is therefore not an instance of morphemic apophony, if it is apophony at all. The same reasoning applies to all cases of suprasegmental, tone or stress, morphemes, of which Spencer (1998) also gives examples.

We see now what makes English irregular plurals (e.g., *foot* / *feet*) and Germanic strong preterites (e.g., English *take* / *took*, German *ich komme* / *ich kam*) real examples of apophonic morphology: there is no exponence of the feature [Past] apart from the alternations /eI/ ~ /U/ and /o/ ~ /a/;⁴⁴ while the alternations are phonologically opaque to varying degrees, they do not lend themselves to an analysis in terms of linking of a floating segment or feature such as Terena /N/. The ablauted plurals of the Celtic languages look very similar to their English equivalents (cf. Irish *fear* / *fir* 'man / men') and may therefore be allotted to apophonic morphology as well. Likewise, Celtic consonant mutations when they represent the only overt signal of the expressed feature, contrary to Fula mutations: cf. Irish *a theach* 'his house' vs. *a teach* 'her house' vs. *a dteach* 'their house', where lenition /t/ → /h/ marks masculine gender of the 3rd person possessor expressed by *a*, no mutation marks feminine gender of the same, and eclipsis /t/ → /d/ indicates there are several possessors of indeterminate gender (see Fife & King 1998).

The vowel alternations one observes between the Perfective and Imperfective forms of Form I active verbs in Arabic must therefore be considered apophonic, as illustrated in the following chart based on Guerssel & Lowenstamm (1990, 1996; also see Kuryłowicz 1973, Chapter II):

(72) k t b 'to write'	D r b 'to hit'	l b s 'to dress'	k b r 'to get big'
a	a	i	u
=> PF <i>katab(a)</i>	<i>Darab(a)</i>	<i>labis(a)</i>	<i>kabur(a)</i>

⁴⁴ Final /-e/ in German *komme* is the exponent of [1SG], not of tense. Its absence in *kam*, however, is significant, so it is not strictly true that apophony is the sole exponent of [Past] in German.

ya-k t b	ya-D r b	ya-l b s	ya-k b r
u	i	a	u

=> IPF *yaktub(u)* *yaDrib(u)* *ya-lbas(u)* *yakbur(u)*

The vowel linked to the second vowel slot is the thematic vowel, the timbre of which is associated, in a not fully transparent fashion, to the aspectual class of the verbalized root (see above).⁴⁵ In each vertical pair the aspect contrast PF / IPF is marked by one step taken along the apophonic path: a → u, a → i, i → a, and u → u (see Ségéral 1995), and nothing else marks it since the vowels between C₁ and C₂ and after C₃ are clearly epenthetic, and the prefix realizes person and number, much like /-e/ in German (but see fn. 44).⁴⁶

Voice marking, in contrast, as in active *katab* / *yaktub* ‘he wrote / writes’ vs. passive *kutib* / *yuktab* ‘it was / is written’ has nothing to do with apophony if the account presented above is anywhere near correct.

8.2. Sites and templates

The second issue deserving more scrutiny than I can devote it here is the relationship of sites, especially internal sites, to templates. Templates are global constraints on the shape and size of concrete lexical items. Most languages probably show some templatic effects – if only because grammatical words and affixes are usually shorter and syllabically simpler than denoting words – but it cannot be denied that the Semitic languages, Arabic in particular, appear strongly templatic when compared with, e.g., the Indo-European languages. Take, for instance, English *baker* and its CA equivalent *xabbaaz*. There is no constraint in English on the shape and size of the base the derivational suffix /-er/ (spelled /-or/ in *tailor*, CA *xayyaaT*) can be attached to (cf. *producer*, *negotiator*, etc.). CA nouns denoting occupations, in contrast, present a fixed template **CaCCaC** which, outwardly at least, fulfils the function of English /-er/. Examples can be multiplied at will. Templatic constraints were mentioned several times in the foregoing study of VN formation.

There are two ways in which templates and the recourse to internal sites are connected. First, as we saw, templates do not exist by themselves, but they arise as a result of internal site activations. For instance, the apparent **CaCCaC** template of Form II (cf. *kassara* ‘he shattered’) follows from the activation of the VDS encompassing C₂. Likewise, *xabbaaz* ‘baker’ or, more accurately, ‘baking person’ can be analysed as in (73), in which ⟨_{3F} x, b, z⟩ is the root roughly meaning ‘bake’ (cf. *xubz* ‘bread’):

(73) {_W ⟨_{3F} x, b, b, ⟨{N}⟩, z⟩ {_{V<II>}}}

CA *xabbaaz* ‘baker’

Geminated C₂ is a carry-over from V<II> which the form includes, thus inheriting verbal force, while the NDS is realized by /A/ in a consonant slot. ‘Nouns of occupation’ thus appear to be kinds of VNs in which N means ‘Agent’ instead of ‘fact/action of’. Again the observed template is but the surface reflection of the CLR.

Second, internal sites must be located. Actually, this is a standard argument from an acquisitional perspective: insofar as site morphology is productive – and it is – language learners must find outward clues that tell them where to look for the internal sites. Templates

⁴⁵ Naturally, this vowel could be given the same type of analysis as with broken plurals, namely as a glide identifying the vowel slot in what could be defined as an inflectional site. I won’t pursue this here.

⁴⁶ Yet, as is well known, the fact it is prefixes rather than suffixes that spell out person and number is also a feature of Imperfective aspect. The Arabic case, like the German case, might thus not be as purely apophonic as English *take* / *took*.

are such a clue. From the observation that word shapes and sizes are uniform for whole families of lexical units sharing a meaning, generalizations can easily be drawn concerning the location of the morphological processes responsible for this uniformity. Connections thus proceed both ways: internal sites yield templates, and templates make internal sites spottable. Languages with few and/or unremarkable templatic effects, such as English, do not prompt learners to look for internal sites. In such languages, the only unmistakably located positions are the beginnings and the ends of words, hence the almost exclusive recourse to marginal sites.

Irregular plurals and strong preterites are indeed exceptions in English. Apropos the latter, it is perhaps no accident that they also exhibit a templatic constraint, namely that (derivation apart as in *upset*) they must show up as C⁺VC⁺ monosyllables. This is apparent in the strong preterites that are not merely apophonic, but also involve the dental Past suffix, e.g. *taught, bought, kept*, etc. In these forms, the suffix cannot assume its underlyingly syllabic shape *-ed* (cf. **kepped*), but it must consist in the lone consonant /t/ (i.e., devoiced /d/), for no obvious reason except obeying the templatic constraint.⁴⁷

Appendix: derived verbs and VNs of Forms XI-XV

Forms XI-XV are kept apart from the other Forms by the rareness of their use as well as by the peculiar ways in which they are formed. I list them below with the associated VNs, using the conventional root ⟨_ʔ f, ʕ, l⟩:

- XI: PF *if^ʕaalla* (CCACC), IPF *yaf^ʕaallu* (CCACC), VN *if^ʕiilaal* (CCICAC)
 XII: PF *if^ʕaw^ʕala* (CCaUCaC), IPF *yaf^ʕaw^ʕalu* (CCaUCaC), VN *if^ʕii^ʕaal* (CCICAC)
 XIII: PF *if^ʕawwala* (CCaUUaC), IPF *yaf^ʕawwilu* (CCaUUiC), VN *if^ʕiwwaal* (CCUUAC)
 XIV: PF *if^ʕanlala* (CCanCaC), IPF *yaf^ʕanlilu* (CCanCiC), VN *if^ʕinlaal* (CCinCaC)
 XV: PF *if^ʕanlaa* (CCanCA), IPF *yaf^ʕanlaa* (CCanCA), VN *if^ʕinlaa'* (CCinCA')

As a I will now show, Forms XI-XV can be accounted for by assuming that the VDS in them is 'larger' than in derived Forms II-X, as it may insert *two* consonant slots rather than one, in addition to possibly encompassing a root consonant. Under this assumption, XI *if^ʕaalla* may be represented as follows:

$$(74) \{w \langle \substack{\text{ʔ} \\ \text{f}, \text{ʕ}, \text{l}} \rangle, \langle \{V_{\langle \text{XI} \rangle} \langle \text{C}, \text{l} \rangle \} \rangle\} \quad \text{CA Form XI } if^{\text{ʕ}}aalla$$

The VDS hosting V_{⟨XI⟩} encompasses two ordered C slots. The first one (from the beginning of the word) is realized as /A/, hence the long vowel, while the second one is occupied by root C₃ which is thus geminated. The fact that C₃, not C₂, is in the VDS is reminiscent of Form IX (see [41]). As in Form IX also, the VN *if^ʕiilaal* is formed from the extended root produced by derivation, namely ⟨_ʔ f, ʕ, C, l, l⟩, in which C is the C slot added by VDS activation. Linking epenthetic /l/ following C₂ /ʕ/ to this slot accounts for the long [ii], while [aa] is the spell-out of /A/ in C filling the NDS exceptionally located after C₃ as in Form IX. This is shown in (75):

$$(75) \{w \langle \substack{\text{ʔ} \\ \text{f}, \text{ʕ}, \text{C}, \text{l} \langle \{N\} \rangle, \text{l}} \rangle, \langle \{V_{\langle \text{XI} \rangle} \} \rangle\} \quad \text{CA Form XI VN } if^{\text{ʕ}}iilaal$$

⁴⁷ I am grateful to Jean Lowenstamm, p.c., for bringing these facts to my attention.

In Form XII *if^caw^cala*, it is C₂ /^c/ that is encompassed by the VDS, next to the two added C slots, the first of which is filled by /U/, the second by the copy of co-site C₂ :

(76) {_W <_Ṛ f, ^c, {V_{<XII>} <^c, C, C>}, l>} CA Form XII *if^caw^cala*

Phonology realizes the vowel slots following site-internal C₂ and its site-internal copy (/^ca...^ca.../). It ignores the vowel slot following intermediate C = /U/, hence the final word-form where /U/ forms a diphthong [aw] with the preceding vowel (/...^caU^ca.../). VN formation proceeds as in Form XI from an extended root <_Ṛ f, ^c, C, ^c, l>, epenthetic /I/ linking to C, as shown in (77):

(77) {_W <_Ṛ f, ^c, C, ^c, {N}, l> {V_{<XII>}} CA Form XII VN *if^cii^caal*

In Form XIII, no root consonant is encompassed by the VDS. The latter adds two C slots after C₂, both realized as /U/:

(78) {_W <_Ṛ f, ^c, {V_{<XIII>} <C, C>}, l>} CA Form XIII *if^cawwala*

Phonology realizes the vowel slots following C₂ and the second added C. The vowel slot separating the two added Cs remains empty (or is automatically deleted). VN formation proceeds normally from the extended root <_Ṛ f, ^c, U, U, l>:

(79) {_W <_Ṛ f, ^c, U, U, {N}, l> {V_{<XIII>}} CA Form XIII VN *if^ciwwaal*

Notice that epenthetic /I/ links to the vowel slot following C₂, hence short [i] contrasting with the long [ii] of Forms XI and XII VNs.

The peculiarity of Form XIV is that, of the two added C slots of the VDS, the first is identified by the infix /n/, while the second copies C₃, which is therefore encompassed:

(80) {_W <_Ṛ f, ^c, {V_{<XIV>} <C, C, l>}, l>} CA Form XIV *if^canlala*

Form XIV is also made special by the fact that C₃ and its copy do not show up as a geminated consonant *[ll], but the vowel slot between them is phonologically realized. If it were not, a phonologically ill-formed sequence */nll/ would result. Given an extended root <_Ṛ f, ^c, n, l, l>, the VN is as expected:

(81) {_W <_Ṛ f, ^c, n, l, {N}, l> {V_{<XIV>}} CA Form XIV VN *if^cinlaal*

Finally, Form XV also involves the infix /n/, but the second added C slot follows encompassed C₃ and it is realized as /A/:

(82) {_W <_Ṛ f, ^c, {V_{<XV>} <C, l, C>}, l>} CA Form XV *if^canlaa*

The extended root is thus <_Ṛ f, ^c, n, l, C>, hence the VN:

(83) {_W <_Ṛ f, ^c, n, l, {N}, C> {V_{<XV>}} CA Form XV VN *if^cinlaa'*

As a result, the VN is distinguished from the finite verb form only by the timbre of the epenthetic vowel following C₂ ([i] vs. [a]) and by the final glottal stop, the default realization of the second added C.

References

- Anderson, S.R. (1992). *A-morphous Morphology*. Cambridge: Cambridge University Press.
- Arnott, D.W. (1970). *The Nominal and Verbal Systems of Fula*. Oxford: Clarendon Press.
- Aronoff, M. (1994). *Morphology by Itself: Stems and Inflectional Classes*. Cambridge, Mass.: MIT Press.
- Asfour, F.-M. (2001). *Theoretische Aspekte der internen Pluralbildung im Klassischen Arabisch*, Diplomarbeit zur Erlangung des Magistergrades, Universität Wien.
- Badawi, E., M.G. Carter & A. Gully (2004). *Modern Written Arabic: A Comprehensive Grammar*. London: Routledge.
- Barth, J. (1904). "Beiträge zur Pluralbildung des Semitischen", *Zeitschrift der deutschen morgenländischen Gesellschaft* 58, 431-446.
- Bat-El, O. (2003). "The fate of the consonantal root and the binyan in Optimality Theory", *Recherches linguistiques de Vincennes* 32, 31-59.
- Beard, R. (1995). *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation*. Albany: State University of New York Press.
- Bohas, G. & J.-P. Guillaume (1984). *Etude des théories des grammairiens arabes. I. Morphologie et phonologie*. Damas : Institut français de Damas.
- Cooper, William S. (1974). *Set Theory and Syntactic Description*. The Hague: Mouton.
- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Fife, J. & G. King. 1998. "Celtic". In A. Spencer & A.M. Zwicky (eds) *The Handbook of Morphology*, 477-499. Oxford: Blackwell.
- Fleisch, Henri (1961). *Traité de philologie arabe, Vol. I : Préliminaires, phonétique, morphologie nominale*. Beyrouth : Imprimerie catholique.
- (1979) *Traité de philologie arabe, Vol. II : Pronoms, morphologie verbale, particules*. Beyrouth : Dar al-Machreq.
- Grismschaw, J. 1990. *Argument Structure*. Cambridge (Mass.) : MIT Press.
- Guerssel, M. & J. Lowenstamm. 1990. The derivational morphology of the Classical Arabic verb. Ms, Université du Québec à Montréal, Université Paris 7.
- 1996. "Ablaut in Classical Arabic Form I Active Verbal Forms." In J. Lecarme, J. Lowenstamm & U. Shlonsky (eds) *Studies in Afroasiatic Grammar*, 123-134. The Hague : Holland Academic Graphics.
- Haiman, J. 1998. "Hua (Papuan)". In A. Spencer & A.M. Zwicky (eds) *The Handbook of Morphology*, 539-562. Oxford: Blackwell.
- Halle, M. (2000). "Distributed Morphology: impoverishment and fission". In J. Lecarme, J. Lowenstamm & U. Shlonsky (eds.) *Research in Afroasiatic Grammar*, 125-149. Amsterdam: Benjamins.
- Halle, M. & A. Marantz, 1993. Distributed Morphology and the pieces of inflection. In K. Hale & S.J. Keyser (eds) *The view from Building 20: Essays in honor of Sylvain Bromberger*, 111-176. Cambridge (Mass.) : MIT Press.
- Harley, H. & R. Noyer (1999). "State-of-the-article: Distributed Morphology", *GLoT* 4.4., 3-9.
- Harris, Z.S. (1951). *Structural Linguistics*. Chicago: The University of Chicago Press.
- Holes, C. (1995). *Modern Arabic: Structures, Functions and Varieties*. London: Longman.

- Jackendoff, Ray (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- Kayne, R.S. (1994). *The Antisymmetry of Syntax*. Cambridge (Mass.) : MIT Press.
- Kihm, A. (2003). “Les pluriels internes de l’arabe : système et conséquences pour l’architecture de la grammaire”, *Recherches linguistiques de Vincennes* 32, 109-155.
- (2004). “Les pluriels internes arabes : un essai d’anatomie”, *Bulletin de la Société de Linguistique de Paris* 99, 311-361.
- (2005a) “Noun class, gender, and the lexicon-syntax-morphology interfaces: a comparative study of Niger-Congo and Romance languages”. In G. Cinque & R.S. Kayne (eds.) *The Oxford Handbook of Comparative Syntax*, 459-512. Oxford: Oxford University Press.
- (2005b) “Old French noun inflection: a Root-and-Site analysis”, poster presented at the 5th *Mediterranean Morphology Meeting*, Fréjus, September 2005 (to appear on *MMM5* website).
- Kuryłowicz, J. (1973). *Studies in Semitic Grammar and Metrics*. London: Curzon Press.
- Larcher, P. (2003). *Le système verbal de l’arabe classique*. Aix-en-Provence: Publications de l’Université de Provence.
- Marantz, A.P. (1997). “No Escape from Syntax : Don’t Try Morphological Analysis in the Privacy of Your Own Lexicon.” In A. Dimitriadis et al., eds., *Proceedings of the 21st Penn Linguistics Colloquium. UPenn Working Papers in Linguistics*, Philadelphia, pp. 201-225.
- Matthews, P.H. (1974). *Morphology: An Introduction to the Theory of Word-Structure*. Cambridge: Cambridge University Press.
- McCarthy, J.J. (1981). “A prosodic theory of nonconcatenative morphology”, *Linguistic Inquiry* 12, 373-418.
- & A.S. Prince. (1990). “Foot and word in prosodic morphology : the Arabic broken plural”, *Natural Language and Linguistic Theory* 8(2), 209-283.
- Plungjan, V.A. (2000). *Obščaja morfologija: vvedenie v problematiku [General Morphology: Introduction to the Issues]*. Moscow: Editorial URSS.
- Pollard, C. & I.A. Sag (1994). *Head-Driven Phrase Structure Grammar*. Chicago: The University of Chicago Press.
- Prunet, J.-F. (1996). “Guttural Vowels”. In G. Hudson (ed.) *Essays on Gurage Language and Culture*, 175-203. Wiesbaden : Otto Harrassowitz.
- (1998). “When vowels function like consonants”, *Phonological Studies* I, 219-226. The Phonological Society of Japan, Tokyo : Kaitakusha.
- Prunet, J.-F., R. Béland & A. Idrissi (2000). “The mental representation of Semitic words”, *Linguistic Inquiry* 31, 609-648.
- Raffelsiefen, R. (1992). “A nonconfigurational approach to morphology”. In M. Aranoff (ed.) *Morphology Now*, 133-162. Albany: State University of New York Press.
- Ratcliffe, R.R. (1998). *The ‘Broken’ Plural Problem in Arabic and Comparative Semitic : Allomorphy and Analogy in Non-Concatenative Morphology*. Amsterdam : John Benjamins.
- Sadock, J.M. (1991). *Autolexical Syntax : A Theory of Parallel Grammatical Representations*. Chicago: The University of Chicago Press.
- Sag, I.S., T. Wasow & E.M. Bender (2003). *Syntactic Theory: A Formal Introduction, 2nd Edition* Stanford : CSLI Publications.
- Ségéral, P. (1995). Une théorie généralisée de l’apophonie, Doctorat, Université Paris 7.
- Spencer, A. (1998). “Morphophonological operations”. In A. Spencer & A.M. Zwicky (eds.) *The Handbook of Morphology*, 123-143. Oxford: Blackwell.
- (2004). “Morphology: an overview of central concepts”. In L. Sadler & A. Spencer (eds.) *Projecting Morphology*, 67-109. Stanford: CSLI Publications.

- Stump, G.T. (2001). *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.
- (2002). “Morphological and syntactic paradigms: arguments for a theory of paradigm linkage”. In G. Booij & J. van Marle (eds.) *Yearbook of Morphology 2001*, 147-180. Dordrecht: Kluwer.
- Van Valin, Robert D. Jr. & Randy J. LaPolla (1997). *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.
- Williams, E. (1981) “On the notions ‘lexically related’ and ‘head of a word’”, *Linguistic Inquiry* 12, 245-274.
- (2004) “Dumping lexicalism”, ms.
- Wright, W. (1896-1898/1991). *A Grammar of the Arabic Language*. Cambridge : Cambridge University Press.