George Anton Kiraz

Syriac Morphology: From a Linguistic Model to a Computational Implementation

> One must know that for the Syriac nouns there are no rigid rules, from which one can learn to form from singular and masculine nouns, plural and feminine ones. On the contrary, one must learn to understand the inflexion of almost all, through means of tradition.
>
> Elia of Ṣobha († 1049)
> *Syriac Grammar*

This paper proposes a linguistically-motivated model of Syriac morphology which can be implemented in a computational environment. My aim is to formalise the derivation of Syriac stems based on the data provided in Nöldeke (1904: §92ff., §160ff.)[1] and to demonstrate how such derivations can be implemented computationally. I shall have little to say about the historical development of Syriac.

Section 1 outlines the conventions used here. Section 2 introduces the structure of Semitic stems. Sections 3 and 4 discuss verb and noun morphology, respectively. Section 5 is devoted to morphotactics. Finally, section 6 outlines the computational modelling of our linguistic description.

1. Conventions and Transliteration

The following convention has been adopted throughout: Morphemes appear in braces, [ ], surface forms in solidi, / /, and phonological segments in square brackets, [ ]. The symbol # indicates stem boundary. The operator '+' indicates concatenation.

The following abbreviations are used: ACT = active, FEM = feminine, MASC = masculine, NPG = number/person/gender, OBJ = object pronominal suffix, PASS = passive, PERF = perfect, REFL = reflexive, SING = singular, VIM = verbal inflexional marker.

The following system is followed in the transcriptions. Parenthesised symbols indicate emphatic consonants.

[1] Full references are given at the end of the article.

Plosives	*Bilabial*	*Dental*	*Velar*	*Glottal*
Voiceless	p	t (ṭ)	k (q)	ʔ
Voiced	b	d	g	

Fricatives	*Labio-dental*	*Inter-dental*	*Alveolar*	*Palato-alveolar*	*Velar*	*Pharyn-geal*	*Glottal*
Voiceless	p̱	ṯ	s (ṣ)	š	ḵ	ħ	h
Voiced	ḇ	ḏ	z		g̱	ʕ	

	Bilabial	*Dental*
Nasals	m	n
Laterals		l
Rolled		r

Semivowels/ Glides	*Bilabial*	*Palatal*
	w	y

Vowels	*Front*	*Central*	*Back*
Close	i		u
Close-mid	e		o
Open-mid			ɔ
Open			a

(pṯɔħɔ̂=a, zqɔpɔ̂=ɔ, rḇɔṣɔ̂=e, ħḇɔṣɔ̂=i, ɪwɔħɔ̂=o, ʕṣɔṣɔ̂=u)

Orthographically, the vowel system indicates vowel quality, rather than quantity. A macron indicates long vowels, e.g. *â* A circumflex indicates that the vowel is followed by *matres lectionis*, e.g. *ā*.

Rukkɔ̄ḵɔ̂ and Qušš̄ɔyɔ̂ are indicated in the usual way, but are not discussed here. (For an introduction on Rukkɔ̄ḵɔ̂ and Qušš̄ɔyɔ̂, see Kiraz (1995).)

The format of phonological rules used here is based on the formalism current in most works on phonology.

(1) Phonological Formalism

Input → Output / LC ___ RC

where

LC = left context

RC = right context

Input represents the segments affected by the rule. The arrow, ⟶, reads 'is realised as' (no historical development is implied). Output represents the result of the rule. The environment bar ___ indicates the position of Input in a stem. The position is determined by the optional contexts. An example appears in (2).

(2) Formalism Example

V → ϕ / C ___ CVC#

Rule (2) states that "a vowel is realised as ϕ (i.e. deleted) when it appears after a consonant C and before a final closed syllable CVC", e.g. /qababab/ → /qabbab/.

2. Stems

The Semitic stem consists of a root morpheme and a vowel melody morpheme (in addition to affixes in some cases), arranged according to a canonical pattern morpheme. Such patterns are called 'CV-skeleta' (McCarthy 1981) for they consist of sequences of Cs (consonants) and Vs (vowels). The terms 'pattern' and 'template' will be used here interchangeably.

CV-skeleta can be further decomposed into syllabic units. Syllables in Semitic are three types as shown in (3).

(3) Syllabic Typology

a. Open light, denoted by CV.
b. Open heavy, denoted by CV̄.
c. Closed heavy, denoted by CVC.

Open light syllables consist of a consonant and a short vowel, e.g. /qa/; open heavy syllables consist of a consonant and a long vowel, e.g. /qā/; and closed heavy syllables consist of two consonants separated by a short vowel, e.g. /qab/.[2] Syllabic boundaries are indicated by a dot, '·', e.g. 'CV·CVC'.

[2] Closed syllables with long vowels, e.g. /qab/, are called ultra-long (Moscati 1969).

3. Verb Morphology

Consider the various verbal forms in (4). The data incorporates a certain degree of abstraction, representing underlying morphological stems and excluding inflexional markers. Phonological processes are required to derive the parenthesised surface forms.

(4) Verbal Measures:

Measure	ACTIVE	PASSIVE: [ʔet]+
1	pʕal	pʕel
2	paʕʕel	paʕʕal
3	ʔapʕel (ʔapʕel)	ʔapaʕl (ʔettapʕal)

Note that, apart from Measure 1, mood is marked by the following vowel melody morphemes: [a-e] for active stems and [a-a] for passive ones. (The symbol '-' indicates that the segments are separated in the surface form.)

Our model argues that all measures, including Measure 1, are derived from the underlying verbal base template in (5).

(5) Verbal Base Template

CV·CVC

The evidence for this template, particularly the existence of the first V, is abundant. Firstly, the proto-Syriac underlying form of the verb is */qaṭal/ 'to kill PERF', */qaṭalaṯ/ 'to kill PERF SING 3rd FEM' etc., by virtue of the softening of the second consonant (e.g. /kṯab/ 'to write').[3] Secondly, an initial consonant cluster would violate syllabic well-formedness in Semitic (Moscati et al. 1969: §10.1). Thirdly, the scarce number of (pre-Syriac) Aramaic verbs which have come down to us in a vocalised manner contain an initial short vowel, e.g. /šaḇaq/ 'to leave' (Mk xv.34). It is clear that these arguments are diachronically motivated. This may seem in contradiction with my earlier statement that 'I shall have little to say about the historical development of Syriac'. However, there is synchronic evidence that justifies a vowel after the first consonant: It is present in (i) Measure 2 (its absence from other measures is phonologically motivated), and (ii) Measure 1 forms when attached to object pronominal suffixes, e.g. /qaṭleh/ 'to kill PERF OBJ', as we shall see soon.

[3] The plosives [b], [g], [d], [k], [p] and [t] become soft (i.e. become fricatives, [ḇ], [g̱], [ḏ], [ḵ], [p̱] and [ṯ], respectively) in postvocalic position.

Before we start looking at the derivation of the various verbal forms, let us take a look at vowel quality in Measure 1. In active stems, the quality of the first vowel in the base template CV·CVC is invariably [a], which is a general Semitic phenomenon (Moscati et al. 1969: §16.2). As we shall see, this is the same [a] in /qaṭleh/ 'to kill PERF OBJ'. The quality of the second vowel, however, is lexically encoded with respect to each individual root as listed in (6); likewise, the quality of the second vowel in imperfective stems is lexically encoded. The second vowel in passive forms of this measure is invariably [e] as in /ʔeṯqṭel/.

(6) Vowel Quality of Measure 1 (cf. Aronoff (1994))

Class	Perfective Vowel	Imperfective Vowel	Example
Default	a	o	qṭal, neqṭol 'to kill'
	a	e	ʕḇaḏ, neʕḇaḏ 'to make'
Stative intransitive	e	a	dḥel, neḏḥal 'to fear'
	e	o	qreḇ, neqroḇ 'to come'
/kpud/	u	a	kpuḏ, neḵpaḏ

The derivation process of triliteral verbs starts by deriving an underlying stem (US) from which the various measures, including Measure 1, are derived. The US is derived from three morphemes: (i) the above verbal base template, (ii) a triconsonantal root, and (iii) a vowel melody. The Cs of the template are associated with root consonants, e.g. [q-ṭ-l] 'notion of killing', and the Vs with vowel melody segments, e.g. [a-e] 'ACT' or [a-a] 'PASS' (or one of the vowel melodies in (6) for Measure 1). This is illustrated schematically in (7) along the lines of McCarthy (1981).

(7) Underlying Stems

```
                               a   e
                               |   |
a.    */qaṭel/ (active)= C V C V C
                         |   |   |
                         q   ṭ   l
```

```
                                 a   a
                                 |   |
b.   */qaṭal/ (passive)=  C V C V C
                          |   |   |
                          q   ṭ   l
```

The rest of this section looks at the derivation of active and passive forms from USs.

3.1. Active Forms

Measure 1 is derived from the US by applying the vowel deletion rule, stated in (8), which has a wide range of applications in all Aramaic idioms, cf. Nöldeke (1904: § 34, § 43).

(8) Vowel Deletion Rule (VDR)

$V \longrightarrow \phi$ / C ___ CVC.

The rule states that a short vowel in an open light syllable is deleted. Applying VDR on the US in (7a) — but with the Measure 1 vowel melody [a-a] — produces Measure 1 as shown in (9).

(9) Measure 1

*/qaṭal/ $\xrightarrow{\text{VDR}}$ /qṭal/ = ⟨ ⟩.

The first vowel in the US, however, is retained with object pronominal suffixes as illustrated in (10).

(10) Measure 1 + OPS

*/qaṭal/ + [eh] $\xrightarrow{+}$ */qa·ṭa·leh/ $\xrightarrow{\text{VDR}}$ /qaṭleh/ = ⟨ ⟩.

It is worth noticing that the VDR applies right-to-left (i.e. starting from the last segment of the string, moving to the first segment). This is why the lexically marked [a] is deleted, and not the first [a].

Skipping Measure 2 for the moment, Measures 3 is derived by prefixing the morpheme [ʔa] to the US and applying VDR as illustrated in (11).

(11) Measures 3

[ʔa] + */qaṭel/ $\xrightarrow{+}$ */ʔa·qa·ṭel/ $\xrightarrow{\text{VDR}}$ /ʔaqṭel/ = ⟨ ⟩.

Finally, Measure 2 is derived by prosodic circumscription (McCarthy 1993), an operation which allows a morphological rule to apply on a prosodically-delimited substring within a stem. This is best explained by an example (12).

(12) Measure 2

a.	Underlying stem	qaṭel
b.	Parsing first syllable	<qa> ṭel
c.	Prefix C	<qa> C+ṭel
d.	Spread ṭ	<qa> ṭṭel
e.	Result	qaṭṭel = ⟨ ⟩

(12a) shows the active US */qaṭel/. We apply a parsing function which extracts the first syllable (i.e. /qa/) from the US; this results in the structure in (12b). Next, we prefix the Measure 2 morpheme — which consists of a consonant C — to the residue /ṭel/ resulting in the structure in (12c). The value of C is determined from the adjacent consonant to the right, i.e. [ṭ], as shown in (12d), i.e. the spreading of [ṭ].

More formally, let S be a string, $PC_{[k,\ op]}(S)$ parses S into a **kernel** k and a residue r, such that $S = k + r$, and performs the operation *op* on r.[4] For example, $PC_{[CV,\ \text{Prefix C}]}$ (qaṭel) parses /qaṭel/ into /qa/ and /ṭel/ and performs the operation 'Prefix C' on /ṭel/. It is worth noting that the first vowel of the base template is retained.

To summarise, all measures are derived from an Underlying Stem whose prosody is described by the pattern CV·CVC. After associating Cs to consonantal roots and Vs to appropriate vowel melodies, Measure 1 is derived by applying the vowel-deletion rule (VDR). Measure 2 is derived by infixing a consonant after the first syllable by prosodic circumscription; the quality of the consonant is determined from the adjacent radical. Measure 3 is derived by prefixing [ʔa] and applying VDR.

A similar analysis can be applied to quadriliteral forms. Note that there is only one quadriliteral measure, viz. CVC·CVC, e.g. /palheḏ/ 'to scatter'.

The stems sapʕel and šaḏpʕel can be handled either following the analysis of Measure 3 or as quadriliteral forms.

[4] This is a simplified version of 'negative' prosodic circumscription, where the domain of the operation is the residue. For more details on prosodic circumscription, see McCarthy (1993).

3.2 Passive Forms

Passive forms are derived from USs in a similar manner, but with the following two differences: (i) applying the respective passive vowel melody, and (ii) prefixing the reflexive morpheme [ʔet]. This is illustrated in (13).

(13) Passive Forms

a. Measure 1: [ʔet] + */qaṭel/ $\xrightarrow{+}$ */ʔeṯ·qa·ṭel/ $\xrightarrow{\text{VDR}}$ /ʔeṯqṭel/ = ⟨ ⟩.

b. Measure 2 [ʔet] + PC[CV, Prefix C] (qaṭal) $\xrightarrow{\text{PC}}$ [ʔet] + */qaṭ·ṭal/ $\xrightarrow{+}$ /ʔeṯqaṭṭal/ = ⟨ ⟩.

c. Measure 3: [ʔet] + [ʔa] + */qaṭal/ $\xrightarrow{+}$ */ʔeṯ·ʔa·qa·ṭal/ $\xrightarrow{\text{VDR}}$ */ʔeṯʔaqṭal/.

Measure 3 requires the assimilation rule stated in (14) — cf. Nöldeke (1904: §36).

(14) [ʔ⟶t]-Assimilation Rule (ASS)

ʔ → t / t ___ V.

The rule states that [ʔ] assimilates into [t] when preceeded by another [t] and followed by a vowel. The derivation of Measure 3 continues in (15).

(15) Measure 3

*/ʔeṯʔaqṭal/ $\xrightarrow{\text{ASS}}$ /ʔettaqṭal/ ⟨ ⟩.

Notice that the [t] of the reflexive prefix is not soft any more after the assimilation process.

3.3 Imperfect Forms

Imperfective stems are prefixed with a morpheme of the form CV (e.g. [ne] 'SING 3rd MASC'). The inflexions of Measure 1 and Measure 2 are straightforward (16).

(16) Measures 1, 2 — Active

a. Measure 1: [ne] + */qaṭol/ $\xrightarrow{+}$ */ne·qa·ṭol/ $\xrightarrow{\text{VDR}}$ /neqṭol/ = ⟨ ⟩.

b. Measure 2 [ne] + */qaṭṭel/ $\xrightarrow{+}$ */ne·qaṭ·ṭel/ $\xrightarrow{\text{VDR}}$ /nqaṭṭel/ = ⟨ ⟩.

The remaining measures require the deletion rule in (17).

(17) [ʔ]-Deletion Rule (DEL)

ʔ → ϕ / C ___ V.

The rule states that the glottal stop [ʔ] is deleted when preceded by a consonant and followed by a vowel. This is exemplified in the derivations of Measure 3 in (18).

(18) Measure 3

a. Active: [ne] + [ʔa] + */qaṭel/ $\xrightarrow{+}$ */ne·ʔa·qa·ṭel/ $\xrightarrow{\text{VDR}}$ */ne·ʔaq·ṭel/ $\xrightarrow{\text{VDR}}$ */n·ʔaq·ṭel/ $\xrightarrow{\text{DEL}}$ /naqṭel/ = ⟨ ⟩.

b. Passive: [ne] + [ʔet] + [ʔa] + */qaṭal/ $\xrightarrow{+}$ */ne·ʔeṯ·ʔa·qa·ṭal/ $\xrightarrow{\text{VDR}}$ */ne·ʔeṯ·ʔaq·ṭal/ $\xrightarrow{\text{VDR}}$ */n·ʔeṯ·ʔaq·ṭal/ $\xrightarrow{\text{DEL}}$ */neṯ·ʔaq·ṭal/ $\xrightarrow{\text{ASS}}$ /nettaqṭal/ = ⟨ ⟩.

Rule DEL is also used in passive inflections of Measures 1 and 2.

4. Noun Morphology

The derivation of the nominal stem here is limited to stems without formatives. Consider the trilateral nominal forms in (19) Nöldeke 1904: §93-125).

(19) Nominal Stems

Pattern		Absolute	Emphatic	Gloss
a.	CV·CVC	mleḵ	malkɔ̂	king
		pleg̱	plegɔ̂	half
		dheḇ	dahḇɔ̂	gold
		qḏoš	quḏšɔ̂	holiness
b.	CV·CV̄C	ħmɔ̄r	ħmɔ̄rɔ̂	ass
		klīl	klīlɔ̂	crown
		lḇūš	lḇūšɔ̂	garment
c.	CVC·CVC	kaddɔ̄ḇ	kaddɔ̄ḇɔ̂	liar
		ṭabbɔ̄ħ	ṭabbɔ̂ħɔ̂	butcher

		quṭṭɔ̄l	quṭṭɔ̄lɔ̂	murdering
		šallīṭ	šallīṭɔ̂	mighty
		settɔ̄r	settɔ̄rɔ̂	covering
		ʔukkɔ̄m	ʔukkɔ̄mɔ̂	black
d.	CV̄·CV̄C	qɔ̄ṭūl	qɔ̄ṭūlɔ̂	murderer
		pɔ̄ṯūr	pɔ̄ṯūrɔ̂	table
e.	CV̄·CVC	rɔ̄ħem	rɔ̄ħmɔ̂	loving
		ʕɔ̄lam	ʕlɔ̄mɔ̂	eternity

Each group in (19) represents stems in a specific pattern. The classification of (19a-b,e] under their respective patterns will become apparent soon.

Synchronically speaking, the view of classical native grammarians that the construct (*gḏɔ̄mɔ̂*, lit. 'syncope') is derived "by the omission of one or two letters from the end of the noun, and the changing [of the position and quality of the] remaining vowel",[5] as Bar Ebrɔ̄yô tells us in his *ṣemḥê*, does not hold in modern morphological theories. Although subtractive morphology (where a morpheme is subtracted from one word to derive another) appears in some languages,[6] subtraction cannot be maintained in (19a,e) because /malkɔ̂/ – [ɔ̂] ≠ /mleḵ/ and /rɔ̄ħmɔ̂/ – [ɔ̂] ≠ /rɔ̄ħem/. Hence, neither the surface absolute can be a derivative of the surface emphatic, nor can the surface emphatic be a derivative of the surface absolute.

Nöldeke (1904: §93) implicitly states that forms like those in (19a) are originally with a short vowel both after the first and the second radicals. The original form can be obtained by collating all the information of the absolute and the emphatic in one string. The operation is illustrated in (20) (without showing the emphatic suffix).

(20) Collating the Absolute and Emphatic Forms

m		l	e	k
m	a	l		k
m	a	l	e	k

[5] Bar Ebrɔ̄yô (1983: p. 67); this notion is also maintained by later traditional native grammarians (Armala (1922), David (1896) and Kfarnisiy (1962)).

[6] For example, in the Muskogean language Koasati, the plural *las* is derived by omitting the rime of the stem, *ap*, from the singular form *lasap* 'to lick' (Martin 1988). For subtractive morphology in Arabic pausal forms, see Hoberman (1995).

The operation produces the US */maleḵ/. This leads to the conclusion that the absolute and emphatic forms in (19a) are derived from an US. The derivation of */maleḵ/ → /mleḵ/ and */maleḵ/ → /mlekɔ̂/ can be achieved by applying VDR (and suffixing the definitive morpheme in the latter).

The forms in (18b) start with a consonant cluster, i.e. CC, which violates syllabic well-formedness in Semitic (Moscati et al. 1969: §10.1). Since the second consonant in such forms is softened (in the case of *bḡɔ̄ḏkpɔ̄ṯ* letters), it is safe to assume an US with a CV on the left periphery. The quality of the vowel, however, can no longer be determined in most cases. In line with traditional grammars, one can mark such vowel with [ə]. Then, the absolute and the emphatic forms in (18b) can be derived by applying VDR which takes care of the deletion of [ə], and the addition of the definitive suffix in case of emphatic forms. The procedure is shown in (21).

(21) CV·CV̄C Forms

a. Absolute: */ħə·mɔ̄r/ $\xrightarrow{\text{VDR}}$ /ħmɔ̄r/.

b. Emphatic: */ħəmɔ̄r/ + [ɔ̂] $\xrightarrow{+}$ */ħə·mɔ̄·rɔ̂/ $\xrightarrow{\text{VDR}}$ /ħmɔ̄rɔ̂/.

The stems in (18c-e) share one characteristic: they are all deverbal nouns. Forms in (18c) tend to be *nomina agentis* (e.g. /kaddɔ̄ḇrɔ̂/), names of occupation (e.g. /ṭabbɔ̄ħɔ̂/), *nomina actionis* (e.g. /quṭṭɔ̄lɔ̂/) or adjectives (e.g. /šallīṭɔ̂/), with few substantive exceptions (e.g. /settɔ̄rɔ̂/) and adjectives of colour (e.g. /ʔukkɔ̄mɔ̂/). Forms in (18d) are similar to the preceeding ones in that they tend to be *nomina agentis*, apart from few substantive forms (e.g. /pɔ̄ṯūrɔ̂/). The same can be said about forms in (18e) which tend to be active participle forms of the Measure 1 verb, again apart from few substantive forms such as /ʕɔ̄lam/; in this case, the VDR applies in the emphatic derivation.

We conclude from the above analysis that nominal stems which start with a heavy syllable are, in most cases, derived from the verb; rare cases can be classified as being atemplatic.

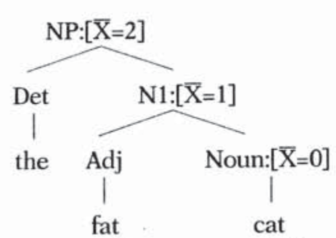
To summarise: Absolute and emphatic forms of stems which start in a light syllable are derived from underlying nominal stems. In the cases of CV·CVC stems, the original vowels are reconstructed by assuming the absolute vowel in place of the second radical and the emphatic vowel in place of the first radical. The first vowel of CV·CV̄C stems, however, cannot be determined and, hence, is indicated by [ə]. The remaining stems are derived from the verb or are atemplatic.

5. Morphotactics

Morphotactics defines the set of licit combinations of morphemes. In verbs, for example, a pattern, a root and a vocalism do not always produce a free stem. Apart from Measure 1, all other verbal measures are bound: they require a 'measure morpheme' which indicates the measure in question, e.g. the prefix [ʔa] in Measure 3. Additionally, passive forms are marked by the 'reflexive morpheme' [ʔet], while active forms are not marked at all.

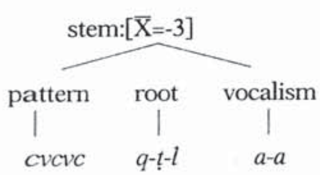
This structure of stems can be handled hierarchically using $\bar{X}$-theory notation. In $\bar{X}$-syntax, lexical categories (e.g. noun, verb and adjective) are assigned $\bar{X}$=0 and the respective phrasal constituents (e.g. noun phrase, verb phrase etc.) are assigned $\bar{X}$=2. Constituents which lie between lexical categories and phrases are assigned $\bar{X}$=1. This is illustrated in (22).

(22) Parsing *the fat cat*



We extend $\bar{X}$-notation as follows. USs are assigned $\bar{X}$=-3 as illustrated in (23).[7]

(23) Parsing the Underlying Stem

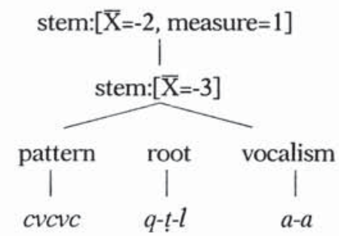


[7] Note that in reality there is no linear precedence between the daughters in (23).

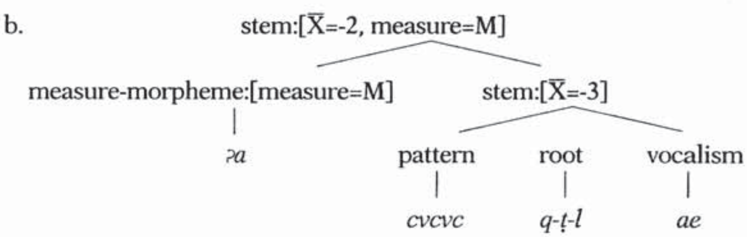
Stems whose measure is known are assigned $\bar{X}$=-2. This is exemplified in (24a) where Measure 1 does not take a measure morpheme, and (24b) where Measure 3 takes the prefix [ʔa]. Note that the value M in (24b) is determined from the prefix.

(24) Stems with Known Measure

a.



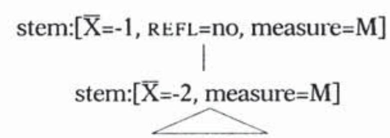
b.



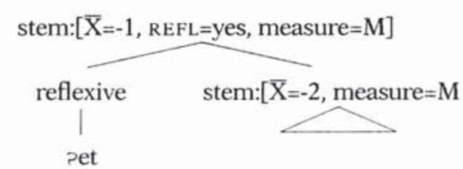
Stems whose reflexivity is known are assigned $\bar{X}$=-1. Active stems do not require the reflexive morpheme (25a), while passive ones do (25b). (The triangle below represents the trees in (24) above.)

(25) Stems with Known Reflexivity

a.



b.



In a similar manner, stems whose inflexion is known are assigned $\bar{X}$=0. For example, the singular 3rd masculine inflexion does not take inflexional markers in the perfect (26a). Other inflections require such markers in the form of prefix, suffix and circumfix. The latter is illustrated in (26). Note that the value of NPG is carried from the inflexional markers.

(26) Parsing of /qṭal/ and /neqṭlūn/

a. stem:[$\bar{X}$=0, REFL=R, measure=M, npg=s3m]
|
stem:[$\bar{X}$=-1, REFL=R, measure=M]

b. stem:[$\bar{X}$=0, REFL=R, measure=M, npg=NPG]

vim:[npg=NPG] stem:[$\bar{X}$=-1, REFL=R, measure=M vim:[npg=NPG]
| |
ne *un*

A similar approach using $\bar{X}$-notation can be used for parsing nominal forms.

6. Computational Illustration

The linguistic model described here was implemented using the SemHe morphological analyser. The technical aspects of the SemHe system have no place here. They are described by Kiraz (1996b) and Kiraz (1996a). Before embarking on our discussion, an introduction to 'two-level morphology' is in order.

6.1. Two-Level Morphology

Computational morphology is a subfield of computational linguistics. It concerns itself with researching applications (computer programs) which can analyse texts. Consider for example a concordance generator. The generator must first identify each word in the text and mark it with its morphological features (e.g. verb PERF SING, etc.). For each word, the program looks in a lexicon (a list of the words in a particular language) and tries to find an identical match. In order for the program to succeed, the lexicon must contain all the forms and inflexions of each word (e.g. book, books, booked, booking) resulting in a huge lexicon. A more efficient search can be achieved by listing stems and morphemes in the lexicon (e.g. book, -s, -ed, -ing) and having a morphological component derive words. Note that the lexicon will have one entry of the suffix -s which would apply to many words.

The most common trend in computational morphology is 'two-level morphology' (Koskenniemi 1983). This model was extended by Kiraz (1994: et seq.) in order to handle Semitic morphology.

Two-level morphology defines two levels of linguistic representations: lexical and surface. The former represents lexical forms (e.g. English [move] and [ed]), while the latter represents surface forms (e.g. English /moved/). Note that the [e] in [move] is deleted on the surface. The [e]-deletion rule takes place when [e] is preceded by a consonant ([v] is this case) and followed by [e] of [ed]. (This is a simplified version of the rule.) Such rules are expressed using the following formalism (Pulman and Hepple 1993).

(27) Two-Level Formalism

LLC	–	LEX	–	RLC	{⇒, ⇐}
LSC	–	SURF	–	RSC	

where

LSC	=	left surface context	LLC	=	left lexical context
SURF	=	surface form	LEX	=	lexical form
RSC	=	right surface context	RLC	=	right lexical context

The special symbol * indicates a context which is always satisfied. The operator ⇒ states that LEX *may* surface as SURF in the given context, while the operator ⇔ adds the condition that when LEX appears in the given context, then the surface description *must* satisfy SURF. The latter caters for obligatory rules. The English [e]-deletion is expressed in rule R1 (28a). To derive /moved/ we also need an identity rule which allows every lexical segment to surface without any change. This is expressed in rule R2.

(28) Grammar for /moved/

a. R1	v	–	e	–	e	⇔
	v	–		–	e	
R2	*	–	X	–	*	⇒
	*	–	X	–	*	

where X is any symbol

b.

m	o	v	e	e	d	*lexical forms*
2	2	2	1	2	2	
m	o	v		e	e	*surface form*

The illustration in (28b) shows the process of the derivation. The top representation gives the lexical forms in question (i.e. [move] and [ed]), while the bottom one gives the surface form /moved/. The numbers indicate the rules in (28a) which sanction the mapping of the segments.

In order to handle Semitic, we use multiple lexical representations as follows: one representation for the pattern, one for the root and one for the vocalism as we shall see below.

The remaining of this section describes the lexicon and the morphological rules required from modelling our description of Syriac.

6.2. *Lexicon*

We assume that the lexicon contains morphemes, the smallest units of morphological analysis. Recall that a Semitic stem consists of a root morpheme and a vowel melody morpheme (in addition to affixes in some cases), arranged according to a canonical pattern morpheme. In order to be able to describe this, we make use of three sublexica: one for patterns, one for roots and one for vowel melodies. Since the segments of patterns correspond to segments in surface forms, we shall use the first sublexicon for prefixes and suffixes as well.

The following lexicon lists the morphemes discussed in section 3.

(29) Lexicon

Sublexicon	Morpheme
1	[cvcvc]
1	[ʔa]
1	[ʔet]
1	[eh]
2	[qṭl]
3	[aa]
3	[ae]
3	[ae]
3	[aa]

The first column indicates the sublexicon containing the morpheme. The second column gives the morpheme.

6.3. *Morphological Rules*

The following two-level grammar models the above description of Syriac. Note that each lexical expression consists of three elements. The first for pattern/prefix/suffix segments, the second for root segments and the third for vowel melody segments. The three expressions are given in angle brackets, ⟨ ⟩, seperated by a comma. The symbol ε (= ϕ in phonological rules above) denotes the empty string. Capitals denote variables: Cs for consonants and Vs for vowels.

(30) Two-Level Grammar

a. R1	*	–	⟨c,C,ε⟩	–	*	⇒
	*	–	C	–	*	
b. R2	*	–	⟨v,ε,V⟩	–	*	⇒
	*	–	V	–	*	
c. R3	*	–	$\langle v,\varepsilon,V_1\rangle$	–	*	⇔
	C_1	–		–	$C_2V_2C_3$	
d. R4	⟨cv,*,*⟩	–	$\langle c,C_1,\varepsilon\rangle$	–	$\langle vc,C_3,V_2\rangle$	⇔
	V_1	–	C_2C_2	–	V_2C_3	
e. R5	*	–	⟨X,ε,ε⟩	–	*	⇒
	*	–	X	–	*	
f. R6	*	–	⟨ʔ,ε,ε⟩	–	*	⇔
	t	–	t	–	V	
g. R7	*	–	⟨ʔ,ε,ε⟩	–	*	⇔
	C	–		–	V	
h. R8	$\langle C_1,*,*\rangle$	–	$\langle V_1,\varepsilon,\varepsilon\rangle$	–	$\langle C_2V_2C_3,*,*\rangle$	⇔
		–		–		

Rule R1 states that the pattern segment [c] and the root segment [C] (i.e. any consonant) — without any vowel melody segment — correspond to the segment [C] on the surface. This is illustrated in the first column in (31a). (The numbers between the surface form and the lexical forms indicate the rules in (30) which sanction the mappings.) In a similar manner, rule R2 states that the pattern segment [v] and the vowel melody

[V] (i.e. any vowel) — without any root segment — correspond to the segment [V] on the surface. This is illustrated in the penultimate column in (31a).

The vowel deletion rule (VDR) is represented by R3. The center of the rule is similar to R2 except that the vowel segment does not appear on the surface (i.e. it is deleted). The contexts ensure that this takes place only in open light syllables. This is illustrated in (31a).

Rule R5 allows any lexical segment (in prefixes and suffixes) to appear on the surface. This is illustrated in the suffix in (30b).

(31) Measure 1

a. /qṭal/

	a		a		*vocalism*
q		ṭ		l	*root*
c	v	c	v	c	*pattern*
1	3	1	2	1	
q		ṭ	a	l	*surface*

b. /qaṭleh/

	a		a			
q		ṭ		l		
c	v	c	v	c	e	h
1	2	1	3	1	5	5
q	a	ṭ		l	e	h

c. /ʔetqṭel/

				a		e	
			q		ṭ		l
ʔ	e	t	c	v	c	v	c
5	5	5	1	3	1	2	1
ʔ	e	t	q		ṭ	e	l

Rule R4 derives Measure 2 by doubling the second consonant. The contexts ensure that this takes place only on the second consonant and in the correct environment as illustrated in (32).

(32) Measure 2

a. /qaṭṭel/

	a		e	
q		ṭ		l
c	v	c	v	c
1	2	4	2	1
q	a	ṭṭ	e	l

c. /ʔetqṭṭal/

				a		e	
			q		ṭ		l
ʔ	e	t	c	v	c	v	c
5	5	5	1	2	4	1	2
ʔ	e	t	q	a	ṭṭ	a	l

The same rules are used in the derivation of Measure 3 as illustrated in (33).

(33) Measure 3 – Active

			a		e	
		q		ṭ		l
ʔ	a	c	v	c	v	c
5	5	1	3	1	2	1
ʔ	a	q		ṭ	e	l

Rule R6 models the [ʔ→ t]-assimilation rule (ASS), while rule R7 models the [ʔ]-deletion rule (DEL). These are illustrated in the reflexive derivation of Measure 3 as shown in (34).

(34) Measure 3 – Passive

a. ʔettaqṭal

						a		a	
					q		ṭ		l
ʔ	e	t	ʔ	a	c	v	c	v	c
5	5	5	6	5	1	3	1	2	1
ʔ	e	t	t	a	q		ṭ	a	l

b. nettaqṭal

								a		a	
							q		ṭ		l
n	e	ʔ	e	t	ʔ	a	c	v	c	v	c
5	8	7	5	5	6	5	1	3	1	2	1
n			e	t	t	a	q		ṭ	e	l

Note that a second version of VDR is modelled in R8. This version deletes a non-stem vowel as shown in (34). A similar approach is used to describe the derivation of nominal forms.

6.4. A SemHe Sample Session

The following sample session was produced using SemHe. The session gives the analysis of /qṭal/ based on the above grammar. The output of the ana (for analyse) command is listed below. ([ṭ] is represented by T.)

```
tli> ana qTal
  Surface Sequence: qTal
  Lexical Sequence: Tape 3: a a
                    Tape 2: q T l
                    Tape 1: c v c v c
  Partitions:
   Rule R1:  Lex: c q _    Surf: q
   Rule R3:  Lex: v _ a    Surf: _
   Rule R1:  Lex: c T _    Surf: T
```

```
10        Rule R2:  Lex: v _ a    Surf: a
11        Rule R1:  Lex: c l _    Surf: l
12       Word Syntax:
13     verb:[bar=0,measure=1,reflexive=no,npg=s3m]
```

Line 1 gives the command line (tli is the prompt for two-level interpreter). Line 2 gives the surface form. Lines 3-5 give the lexical sequences: pattern, root and vocalism (the term 'tape' is computer jargon). Lines 7-11 give the lexical-surface mapping according to the above morphological rules. For example, line 7 states that by rule R2, lexical [c], [q] and ε (the symbol "-" above) map to [q] on the surface. Lines 13 give the morphosyntactic features of the word based on some morphosyntactic grammar.

7. Future Work

This paper gave a framework under which computational descriptions of Syriac morphology can be based. The above discussion is by no means exhaustive. Nominal stems with formatives and deverbal stems have not been discussed fully, and none of the cases of derivational morphology have been mentioned.

Additionally, various issues which arise in writing two-level grammars of Syriac were not described here. Some of these issues can be found in Kiraz (1996a), while others (e.g. the morphosyntactic handling of the *bdul* prefixes, agreement of object pronominal suffixes and stems, etc.) await further investigation.

References

Armala, I. (1922). *al-'uṣūl al-'ibtida'iyyah fī al-lughah al-suryaniyyah*. Jesuit Fathers, Beyrut.

Aronoff, M. (1994). *Morphology by Itself. Stems and Inflectional Classes*. Linguistic Inquiry Monographs 22. The MIT Press.

Bar Ebrōyô (1983). *ktābâ d-ṣemḥê*. Syrianska Riksförbundets Kultur Kommitté, Sweden, 2 edition.

David, C. (1896). *kitâb 'al-lum'ah al-šahiyyah fī naḥw 'al-luğah 'al-suryâniyyah*. Dominican Press, Mossoul, 2 edition.

Hobermann, R. (1995). "Subtractive morphology and morpheme identity in Arabic pausal forms". In Booij, G. and van Marle, J., (eds), *Yearbook of Morphology* 1995, pp. 161-74. Kluwer Academic Publishers.

Kfarnīsıy, B. (1962). *ğrāmaṭīq 'al-luğah 'al-'ārāmiyyah 'al-suryâniyyah*. Lebanon, 2 edition.

Kiraz, G. (1994). "Multi-tape two-level morphology: a case study in Semitic non-leaner morphology". In *COLING-94: Papers Presented to the 15th International Conference on Computational Linguistics*, vol. 1, pp. 180-6.

Kiraz, G. (1995). *Introduction to Syriac Spirantization*. Bar Hebreaus Verlag, The Netherlands.

Kiraz (1996a). *Computational Approach to Non-Linear Morphology*. PhD thesis, University Cambridge.

Kiraz, G. (1996b). "ṢEMḤE: A generalised two-level system". In *Proceedings of 34th Annual Meeting of the Association for Computational Linguistics*.

Koskenniemi, K. (1983). Two-Level Morphology. PhD thesis, University of Helsinki.

Martin, J. (1988). "Subtractive morphology as dissociation". In *Proceedings of the Seventh West Coast Conference on Formal Linguistics*. Stanford Linguistics Association.

McCarthy, J. (1981). "A prosodic theory of nonconcatenative morphology". *Linguistic Inquiry*, 12 (3), pp. 373-418.

McCarthy, J. (1993). "Template form in prosodic morphology". In Stvan, L. et al., (eds), *Papers from the Third Annual Formal Linguistics Society of Midamerica Conference*, pp. 187-218. Indiana University Linguistics Club, Bloomington.

Moscati, S. Spitaler, A., Ullendorff, E., and von Soden, W. (1969). *An Introduction to the Comparative Grammar of the Semitic Languages: Phonology and Morphology*. Porta Linguarum Orientalium. Otto Harrassowitz, Wiesbaden, 2 edition.

Nöldeke, T. (1904). *Compendious Syriac Grammar*. London. Translated by J. Crichton.

Pulman, S. and Hepple, M. (1993). "A feature-based formalism for two-level phonology: a description and implementation". *Computer Speech and Language*, 7 pp. 333-58.

Bell Laboratories
Room 2D-513
700 Mountain Ave
Murray Hill NJ 07974
U.S.A.

George Anton Kiraz