



2. Lemmas

Lemmas represent words as they appear in dictionaries. Our system of Syriac lemmatisation is based on Sokoloff’s *Syriac lexicon*<sup>4</sup> used as “reference dictionary”:

1. Lemmas correspond to the headwords found in Sokoloff’s dictionary with eastern vocalization, all diacritical marks (including spirantization<sup>5</sup> and linea occultans), in the emphatic state<sup>6</sup>:
  - a. Form ܠܘܘܘܢܐ: lemma ܠܘܘܘܢܐ (“love” Sokoloff, p. 419)
  - b. Form ܠܘܘܘܢܐ: lemma ܠܘܘܘܢܐ (“new” Sokoloff, p. 418)
2. Lemmas of the verbs should also be vocalised with eastern vocalisation. However, the vocalisation is only specified by Sokoloff for the V1 (*pāʿal*) form in brackets. The lemmatiser will then follow Payne-Smith’s vocalisation by adapting it one-to-one to East Syriac (see conversion table below). The diphthong *aw* ܐܘ in West Syriac corresponds to *āw* ܐܘ in East Syriac (see Nöldeke §49). In case of doubt regarding the vocalisation of verbal forms, the lemmatiser should consult Toma Audo’s East Syriac dictionary.
  - a. Form ܠܘܘܘܢܐ: lemma ܠܘܘܘܢܐ (“to return” Sokoloff, p. 350; J. Payne-Smith, p. 105 ܠܘܘܢܐ)
  - b. Form ܠܘܘܘܢܐ: lemma ܠܘܘܘܢܐ („to add” Sokoloff, p. 577; J. Payne-Smith, p. 194 ܠܘܘܢܐ; Audo, p. 435)
  - c. Form ܠܘܘܘܢܐ: lemma ܠܘܘܘܢܐ (“to be added” Sokoloff, p. 577; J. Payne-Smith, p. 194 ܠܘܘܢܐ; Audo, p. 435)

V1	ܠܘܘܢܐ	ܠܘܘܢܐ
V2	ܠܘܘܢܐ	ܠܘܘܢܐ
V3	ܠܘܘܢܐ	ܠܘܘܢܐ
V4	ܠܘܘܢܐ	ܠܘܘܢܐ
V5	ܠܘܘܢܐ	ܠܘܘܢܐ
V6	ܠܘܘܢܐ	ܠܘܘܢܐ

3. Lemmas of proper names (anthroponyms and toponyms) follow the spelling found in Payne Smith’s *Thesaurus*<sup>7</sup>, without vocalization:
  - a. Form ܘܫܘܥܐ: lemma ܘܫܘܥܐ (“Jesus” Payne Smith, col. 1638)
  - b. Form ܘܫܘܥܐ: lemma ܘܫܘܥܐ (“Dadišo” Payne Smith, col. 824)
4. Homographic lemmas are distinguished by a number conforming to the usage of Sokoloff’s dictionary:
  - a. Form ܘܫܘܥܐ: lemma “1 ܘܫܘܥܐ” (“to hope” Sokoloff, p. 964)
  - b. Form ܘܫܘܥܐ: lemma “2 ܘܫܘܥܐ” (“to announce” Sokoloff, p. 965)
5. Word unrecorded neither in the reference dictionary nor in the database of the GREgORI project is recorded with lemma as intuitively expected by linguists:
  - a. Form ܠܘܘܘܢܐ: lemma ܠܘܘܘܢܐ “being recompensed”
  - b. Form ܠܘܘܘܢܐ: lemma ܠܘܘܘܢܐ “prone to disease”

<sup>4</sup> M. SOKOLOFF, *A Syriac Lexicon: A Translation from the Latin, Correction, Expansion, and Update of C. Brockelmann’s Lexicon Syriacum*, Piscataway, 2009 (2d corrected print, 2012).

<sup>5</sup> More information about spirantisation can be found in Appendix

<sup>6</sup> For the sake of readability, we will only present from this point forward monolexical occurrences. Passive participles in the absolute state in Sokoloff will still receive a lemma in the emphatic state.

<sup>7</sup> R. PAYNE SMITH, *Thesaurus syriacus*, Oxford, 1871-1901 (reprint Hildesheim, 2006).

### 3. POS-tags

Each lemma receives a tag that indicates the part of speech, the morphosyntactic category, to which it belongs. This tag characterizes the lemma at the lexical level in general. It does not describe in any way the particular use of the lemma in a specific context. POS-tags used in our system of analysis are as follows:

- ADJ:** adjective (ܡܘܨܝܩܐ “lacking” Sokoloff, p. 1024)
- ADV:** adverb (ܐܘܕܢܐ “again” Sokoloff, p. 1626)
- CARD:** cardinal number (ܘܚܕܐ “one” Sokoloff, p. 413)
- INTJ:** interjection (ܘܘܐ “woe!” Sokoloff, p. 357)
- NAME:** proper noun
  - NAME\_ant:** anthroponym ܐܕܡܐ “Adam” (Payne Smith. col. 38)
  - NAME\_top:** toponym ܡܘܨܝܩܐ “Egypt” (Payne Smith. col. 2196)
- NOUN:** noun (ܐܚܘܝܬܐ “brother” Sokoloff, p. 25)
- ORD:** ordinal number (ܦܪܘܫܬܐ “first” Sokoloff, p. 1319)
- PART:** particle<sup>8</sup> (ܘܐ “and” Sokoloff, p. 357)
- PRO\_dem:** demonstrative pronoun (ܗܝܘܐ “this one” Sokoloff, p. 346)
- PRO\_ind:** indefinite pronoun (ܡܝܘܨܝܩܐ “something” Sokoloff, p. 715)
- PRO\_int:** interrogative pronoun (ܡܘܨܝܩܐ “how” Sokoloff, p. 34)
- PRO\_pers:** personal pronoun (ܐܝܢܐ “I” Sokoloff, p. 58)
- V1–V33:** verbs<sup>9</sup>
  - V1:** ܡܘܨܝܩܐ “to confine” (Sokoloff, p. 411)
  - V3:** ܘܘܐܢܘܢܐ “to announce” (Sokoloff, p. 965)
  - V5:** ܡܘܨܝܩܐ “to terrify” (Sokoloff, p. 374)
  - ...

### 4. Roots

At the level of the root, two categories are distinguished.

1. **Base Nouns** (primary nouns)<sup>10</sup> (vocalized, if the vocalization is attested):
  - a. Lemmas: ܐܕܡܐ, ܐܚܘܝܬܐ and ܐܕܡܐܘܕܢܐ: root ܐܕܡܐ (Sokoloff, p. 1)
  - b. Lemmas: ܡܘܨܝܩܐ, ܘܘܐܢܘܢܐ: root ܡܘܨܝܩܐ (Sokoloff, p. 965)
2. **Consonantal Roots** (not vocalized), linked to a base verb from which nouns and adjectives are derived:
  - a. Lemmas: ܐܘܪܐܝܬܐ, ܐܘܪܐܝܬܐ, ܐܘܪܐܝܬܐܘܕܢܐ, ܐܘܪܐܝܬܐܘܕܢܐ, ܐܘܪܐܝܬܐܘܕܢܐ: root ܐܘܪܐܝܬܐ (Sokoloff, p. 374)
  - b. Lemmas: ܡܘܨܝܩܐ, ܡܘܨܝܩܐ, ܡܘܨܝܩܐܘܕܢܐ: root ܡܘܨܝܩܐ (Sokoloff, p. 1352)

Particular cases:

1. Homographic roots are distinguished by a number conforming to the usage of Sokoloff’s dictionary:
  - a. Lemma 1 ܘܘܐܢܘܢܐ: root 1 ܘܘܐܢܘܢܐ
  - lemma 2 ܘܘܐܢܘܢܐ: root 2 ܘܘܐܢܘܢܐ
  - lemma 3 ܘܘܐܢܘܢܐ: root 3 ܘܘܐܢܘܢܐ
  - lemma 4 ܘܘܐܢܘܢܐ: root 4 ܘܘܐܢܘܢܐ
  - (as Sokoloff, pp. 1133-1134)

<sup>8</sup> Particles are tagged as conjunction and preposition in Sokoloff’s *Syriac Lexicon*.

<sup>9</sup> Complete table of the verbal POS-tags can be found at the following address: [https://www.gregoriproject.com/pdf/POS\\_SYC.pdf](https://www.gregoriproject.com/pdf/POS_SYC.pdf).

<sup>10</sup> In this case, it is the noun or the adjective that is original. The majority of base nouns falls into the following categories: family (father, brother, etc.), body parts (heart, foot, etc.), animals (bull, rabbit, etc.), plants (herbs, garlic, etc.), stones and metals (clay, gold, etc.), temporal concepts (year, eternity, etc.), geographical concepts (earth, rivers, etc.), architectural concepts (dike, roof, etc.), tools (basket, vase, etc.) and foreign words (palace, paradise, etc.). They can themselves be at the origin of verbs, which are qualified as *denominative verbs*, generally used in a derived *form*. Taking into account the base nouns obviates the need to create artificial verbal roots that are supposed to account for them (see B. KINDT, J.-Cl. HAELEWYCK, A.B. SCHMIDT, N. ATAS, *La concordance bilingue grecque-syriaque des Discours de Grégoire de Nazianze*, in *BABELAO*, 7 (2018), p. 63-64 ; document freely available [here](#)).

- b. Lemmas ܐܘܢܐ, ܐܘܢܐܘܢܐ, ܐܘܢܐܘܢܐܘܢܐ and ܐܘܢܐܘܢܐܘܢܐܘܢܐ: root ܐܘܢ; lemmas ܐܘܢܐܘܢܐ and ܐܘܢܐܘܢܐܘܢܐ: root ܐܘܢ (as Sokoloff, pp. 1044-1045)
- 2. If the word is not recorded in reference dictionaries:
  - a. If the root (consonantal or base noun) is known, this root is used
    - i. Lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐ “being recompensed”: root ܐܘܢ
    - ii. Lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐ “prone to sickness”: root ܐܘܢ
  - b. If the root remains unknown, lemma is used as root
    - i. Lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐ: root (base noun) ܐܘܢܐܘܢܐܘܢܐܘܢܐ
    - ii. Lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐ: root (base noun) ܐܘܢܐܘܢܐܘܢܐܘܢܐ
    - iii. Lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐ: root (base noun) ܐܘܢܐܘܢܐܘܢܐܘܢܐ

## 5. Origin

Each root receives an indication specifying if the word is borrowed or not from another language than Syriac. In our database, Syriac words are characterized by the sign “0” and non-Syriac word by the sign “-1” (e.g. ܐܘܢܐܘܢܐ, from Greek ἄνθρωπος). In our concordances in PDF format, non-Syriac roots are therefore automatically characterized by the sign “Ⓢ” (for “loanword”). Proper names are not taken into account and are conventionally characterized by the “0” sign.

## Appendix 1: Spirantisation

1. The *rukkākhā* and *quššāyā* points are to be indicated when spirantisation or non-spirantisation are not expected. Spirantisation is expected after a vowel or after two consonants one after another at the beginning of a syllable, and non-spirantisation after a consonant. The points are already indicated by Sokoloff, except for the verbs.
  - a. Form ܐܘܢܐܘܢܐܘܢܐܘܢܐ: lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐ (“milk” Sokoloff, p. 452)
  - b. Form ܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐ: lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐ (“to receive” Sokoloff, p. 1311)
  - c. Form ܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐ: lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐ (“to send” Sokoloff, p. 1514)
2. In the case of non-spirantisation due to assimilation, both letters are to be marked (especially in verbal forms with prefix -ܐܘܢܐܘܢܐܘܢܐ).
  - a. Form ܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐ: lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐ (“to be added” Sokoloff, p. 577)
  - b. Form ܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐ: lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐ (“to be amazed at” Sokoloff, p. 311)
  - c. Form ܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐ: lemma ܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐܘܢܐ (“to be destroyed” Sokoloff, p. 535)

Appendix 2: Keyboard Shortcuts

	Belgian Keyboard	German Keyboard	English Keyboard
<b>Vowels</b>			
a ܐ (ܐ)	SHIFT + W	SHIFT + Y	SHIFT + Z
ā ܐ̄ (ܐ̄)	SHIFT + X	SHIFT + X	SHIFT + X
e ܐ (ܐ)	SHIFT + C	SHIFT + C	SHIFT + C
ī/ū ܐ̄ (ܐ̄/ܐ̄)	SHIFT + V	SHIFT + V	SHIFT + V
o ܐ (ܐ)	SHIFT + B	SHIFT + B	SHIFT + B
ē ܐ̄ (ܐ̄)	SHIFT + N	SHIFT + N	SHIFT + N
<b>Spirantisation</b>			
<i>Rukkākhā</i> ܐ (ܐ)	SHIFT + ,	SHIFT + M	SHIFT + M
<i>Quššāyā</i> ܐ (ܐ)	SHIFT + U	SHIFT + U	SHIFT + U
<b>Lengthening Line</b> <i>(kashīda)</i> <sup>11</sup> : - (ܐ̄)			
	SHIFT + J	SHIFT + J	SHIFT + J
<b>Plural Marker</b> <i>(syāmē)</i> : .. (ܐ̄)			
	SHIFT + I	SHIFT + I	SHIFT + I
<b>Linea occultans (<i>mbaṭṭlānā</i>):</b>			
Below - (ܐ̄)	SHIFT + L	SHIFT + L	SHIFT + L
Above - (ܐ̄)	SHIFT + O	SHIFT + O	SHIFT + O
<b>Syriac abbreviation mark:</b> ܐ̄	<sup>2</sup> (before the word)	^ (before the word)	` (before the word)

<sup>11</sup> In order to distinguish single letter particles ܐ, ܐ̄, ܐ̄, ܐ̄ from numbers, particles are followed by the kashida ܐ̄, ܐ̄, ܐ̄, ܐ̄. In accordance with Sokoloff's dictionary (p. 295), we add also the kashida after the possessive particle -ܐ̄ܐ̄. Higher numbers consisting of several letters should not be divided during the lemmatisation. E.g., the form ܐ̄ܐ̄ܐ̄ (5550) will be lemmatised as ܐ̄ܐ̄ܐ̄ and not as ܐ̄ܐ̄ܐ̄ܐ̄.