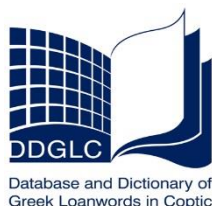


Comprehensive Coptic Lexicon V 1.2



Release Notes

July 22, 2020

Publishers:	Strukturen und Transformationen des Wortschatzes der ägyptischen Sprache (BBAW) Database and Dictionary of Greek Loanwords in Coptic (FU Berlin)
Distributor	Freie Universität Berlin/Refubium
Raw data:	http://dx.doi.org/10.17169/refubium-27566
Processed data:	http://coptic-dictionary.org
Compatibility:	TEI Dictionary Module, TLA XML Schema
Availability:	CC BY-SA 4.0
Change log:	https://github.com/KELLIA/dictionary/wiki/CCL-V-1.2-Changelog

1. Description:

Comprehensive Coptic Lexicon V 1.2 is a joint release of the “Database and Dictionary of Greek Loanwords in Coptic” (Freie Universität Berlin) and the research project “Strukturen und Transformationen des Wortschatzes der ägyptischen Sprache” Thesaurus Linguae Aegyptiae” (Berlin-Brandenburgische Akademie der Wissenschaften). It includes Version 4 of the BBAW lexicon of Coptic Egyptian and Version 2 of the DDGLC lexicon of Greek loan words in Coptic. The major changes and new features are:

- Unified the use of the parenthesis “()” in word forms: they stand now for collocated prepositions and adverbs only, e.g. C38 $\lambda\epsilon$ (εγραί). All other uses were disambiguated into separate forms, e.g.
C84: $\lambda\mu\eta(\epsilon)\iota\eta\epsilon \rightarrow \lambda\mu\eta\iota\eta\epsilon, \lambda\mu\eta\epsilon\iota\eta\epsilon;$
C836: $\epsilon\iota ((\text{N})\text{N}\lambda\epsilon\text{P}\text{N}-) \rightarrow \epsilon\iota (\text{N}\lambda\epsilon\text{P}\text{N}-), \epsilon\iota (\text{N}\text{N}\lambda\epsilon\text{P}\text{N}-);$
C920: $\epsilon\iota\eta\epsilon (\epsilon\text{P}\lambda\epsilon\text{OY}, \epsilon\text{I}\text{P}\lambda\epsilon\text{OY}) \rightarrow \epsilon\iota\eta\epsilon (\epsilon\text{P}\lambda\epsilon\text{OY}), \epsilon\iota\eta\epsilon (\epsilon\text{I}\text{P}\lambda\epsilon\text{OY});$
C1466: $(\kappa\omega\tau\alpha) \rightarrow *κ\omega\tau\alpha.$

The old forms obtained the attribute *change="#deprecated #v1.2"*, the new ones are marked with *change="#added #v1.2"* and are linked via the “*corresp= TLA_ID*” attribute. For details please refer to the description of the project: <https://github.com/KELLIA/dictionary/wiki/A0:-Project:-Remove-brackets-from-orthographic-variants>.

- Optimized the data structure:
 - Unified the location of the written forms, which are now in <orth> tag only (#85). Parsed for empty <orth> tags (#109) and tags containing white spaces at word’s beginning (#83), correcting them.
 - Assigned dialect information to Sahidic forms, which, perceived as default, did not have dialect explicitly encoded: <usg type="geo">S</usg>.

- In line with LEX-0 conventions improved the structure of <sense> element, which now contains a unique ID and a single <cit type="translation"> tag. This facilitates the linking of CCL to the databases of semantic relations such as Coptic WordNet.
 - Adopted XML Schema to allow deprecated forms and entries as well as versioning.
- Resolved 45 issues, which included correcting the orthographic, grammatical and semantic information of the existing entries. E.g. the spelling of bound forms is now uniform: status nominalis words use single hyphen "-" (U+002D: HYPHEN-MINUS), status pronominalis forms utilize double oblique hyphen "≐" (U+2E17: DOUBLE OBLIQUE HYPHEN) (#95). For the full list of resolved bugs see the list in section §2 below.
 - Created new entries:
 - 1st person singular (ⲓ, ⲧ), plural (ⲛ) and 3rd person singular (ⲙ) pronouns (#105);
 - ⲉ- as conditional (#112);
 - ⲗⲟⲓⲉ "a cause", ⲗⲟⲓⲉ "to lie hid" (#116).
 - Improved usability of the Graeco-Coptic section due to exclusion or change of a number of senses:
 - because they are no longer in the DDGLC database¹;
 - because they are in a duplicate entry²;
 - because they represent DDGLC usages that are based on reconstructed or uncertain attestations (so called "0-usages", 295 cases); if the excluded senses were the sole sense for the entry, the entry itself is marked as 'change="deprecated"'³;
 - because the sense description in the CCL equals other sense descriptions in the same entry (e.g. C9211) and thereby appear to be duplicates, but are distinguished not just by DDGLC usage ID but also additional field information in the DDGLC database which is not represented in the CCL⁴;
 - because sense description changed⁵.
 - Linked Greek words (head words of Greek-Coptic lexicon) to Perseus Greek morphology tool via the UI of Coptic Dictionary Online, resolving a related issue: polytonic Greek letters ἄ, ἕ, ἦ, ἱ, ῶ, ῦ und ῷ are now uniformly coded with the regular (modern) Greek tonus characters for the acute (#125).
 - DDGLC lemma IDs are now displayed in entry view of Coptic Dictionary Online.
 - Unique IDs assigned to all newly created or disambiguated forms and entries.

¹ Ref type="coptic_usage::cu_ID": 1073, 1243, 1665, 1977, 2108, 2321, 2720, 2978, 3414, 3837, 4199, 4289, 5136, 5246, 5748, 5765, 5987, 6097, 6212, 7080, 7271, 7411, 8105, 9388, 9730, 10109, 10332, 10477, 10496, 10613, 10653, 10792, 10905, 11662, 11683, 15463.

² Ref type="coptic_usage::cu_ID": 5186, 262, 5192, 5290, 10213, 5184, 7446, 5182, 5193, 5185, 5183, 9283, 4202, 5296, 11090, 2249, 5295.

³ TLA lemma nos. C8218, C8337, C8520, C8723, C8833, C8879, C9093, C9386, C9472, C9947, C10078, C10389, C10539, C10612, C10752, C10889, C10912, C10924, C11055, C11261, C11270.

⁴ Ref type="coptic_usage::cu_ID": 10921, 6149, 11436, 2744, 490, 7573, 2525.

⁵ Ref type="coptic_usage::cu_ID": 5284, 9036, 4170, 5690, 5746, 5831.

Comprehensive Coptic Lexicon V 1.2 contains:

- BBAW lexicon of Coptic Egyptian:
 - Entries without deprecated ones: **8055**
 - Entries total: **8055**
 - Forms without deprecated ones: **18590**
 - Forms total: **18971**
 - Senses total: **8945**

- DDGLC lexicon of Greek loan words in Coptic:
 - Entries without deprecated ones: **3208**
 - Entries total: **3229**
 - Forms without deprecated ones: **13257**
 - Forms total: **13259**
 - Senses total: **6978**

- Comprehensive Coptic Lexicon V 1.2 (BBAW + DDGLC combined):
 - Entries without deprecated ones: **11263**
 - Entries total: **11284**
 - Forms without deprecated ones: **31847**
 - Forms total: **32230**
 - Senses total: **15923**

2. List of resolved bugs

#	Bug (https://github.com/KELLIA/dictionary/issues)	Reported by
#125	oxia vs tonus in polytonic Greek letters	amir-zeldes
#122	Wrong Demotic TLA IDs in the etymological reference	phoenix-mossimo
#76	ΠΟΝΗΡΟΝ contains forms of ΠΟΝΗΡΟΣ	phoenix-mossimo
#84	χρεία: Subst. and Vb. swapped	m-linssen
#121	lemma of thesaurus	amir-zeldes
#109	Empty orthographic values with a number of forms	phoenix-mossimo
#95	Microsoft autocorrect en dashes in lemmas	m-linssen
#103	Add "double oblique hyphen" ("≐") to the virtual keyboard	phoenix-mossimo
#116	Lexicographic errors	phoenix-mossimo
#79	Cosmetic: partinle (sic)	m-linssen
#119	Linking entries to Perseus	phoenix-mossimo
#115	Add DDGLC IDs in entry view	amir-zeldes
#114	Sort forms	phoenix-mossimo

#117	DH Awards 2019 Logo on the title page	phoenix-mossimo
#100	Support nested gramGrp	amir-zeldes
#113	Relative converter form displayed as ⲙⲡⲁⲣⲉ- instead of ⲉ- (C617)	phoenix-mossimo
#112	Entry missing for ⲉ- meaning "if"	amir-zeldes
#104	ⲡⲉ has incorrect cf. and a typo	m-linssen
#105	Suffixpronomen inconsistency and errors	m-linssen
#107	ⲃⲔ ⲛⲉⲗⲟⲟⲗⲉ male?	m-linssen
#108	Parentheses get eaten by search	m-linssen
#110	Typo: from vs form	m-linssen
#111	Inconsistency in the ⲛⲧⲉ- forms	m-linssen
#106	What's the idea on pronouns?	m-linssen
#102	Spellings with supralinear strokes are not searchable	phoenix-mossimo
#82	ⲟⲩⲉ is marked as Subst. instead of Zahlwort	m-linssen
#88	ⲉⲧⲣⲟⲥ - superlinear strokes (only) in TLA lemma no. C10535 < and more	m-linssen
#101	Consistency in punctuation with multiple meanings	phoenix-mossimo
#87	Second meaning of ⲭⲬⲟⲡⲬⲡ	m-linssen
#97	Broken pointers and inconsistencies across references	m-linssen
#86	ⲭⲁ-, contains a comma	m-linssen
#89	ⲣⲉⲩⲧⲭⲁⲛ translation contains epsilon	m-linssen
#92	ⲟⲩⲟⲧⲉ (flour / dough) is male, not female	m-linssen
#85	Put all forms in <orth> element; strip white spaces	phoenix-mossimo
#96	Incomplete cross-references across lemmas	m-linssen
#30	Entries with parentheses don't show up in searches	ctschoeder
#94	ⲡⲟⲟⲩ (ⲡⲣⲟⲟⲩ) even more confusing	m-linssen
#98	ⲧⲁⲟⲩⲉⲟⲩⲟ has two separate entries	m-linssen
#99	ⲟⲩⲡⲛⲣⲉ two entries, odd ANNIS	m-linssen
#93	ⲟⲩⲟⲗⲙ - split across 5 lemmas	m-linssen
#90	ⲁⲙⲁⲣⲧⲉ intr. and tr.? Odd / completely absent entries	m-linssen
#91	Circular reference for https://coptic-dictionary.org/results.cgi?coptic=ⲙⲡ	m-linssen
#78	ⲟⲩⲡⲛⲣⲉ points to noun twice, not the verb	m-linssen
#83	"ⲩⲓ- " and ⲟⲩⲟⲛ as Vb.	m-linssen
#80	ⲛⲟⲖⲛⲉⲖ & ⲛⲟⲖⲛⲟⲖ	ctschoeder
#77	Dialect search crashes	amir-zeldes

3. Known issues and research perspectives

- DDGLC dataset: deprecated entry C8879 contained a wrongly applied sense (ref type="coptic_usage::cu_ID": 4047): it is currently not known or certain that a similar case of a wrongly applied sense is present in the Graeco-Coptic section of the CCL. A systematic check will be applied for a following release.
- BBAW dataset: the grammatical information of “portmanteau” or “layered” forms, which contain two grammatical categories (possessive prefix, designating the possessed item, and possessive suffix, designating the possessor, e.g. ⲛⲁ- (C2353), ⲛⲟⲩ- (C2388), ⲡⲁ- (C2784), ⲡⲟⲩ- (C2787), ⲧⲁ- (C4005), ⲧⲟⲩ- (C11281))), was disambiguated in V1.2. For now a

temporary solution has been chosen – to relegate some of the grammatical information to the definition text in tag <sense>. The preferred solution is to bring the grammatical encoding in accordance with LEX-0:

<https://github.com/KELLIA/dictionary/wiki/Encoding-possessor-%26-possessum-information-with-possessive-pronouns%2C-articles-and-prefixes>.

- The next update will provide a major alignment with LEX-0. For the description of the project and current tasks see: <https://github.com/KELLIA/dictionary/projects/4>

4. Files to be released:

File name	File size (MB)	Modification date
BBAW_Lexicon_of_Coptic_Egyptian-v4-2020.xml	6,41	21.07.2020 18:25
Comprehensive_Coptic_Lexicon-v1.2-2020.xml	11,7	22.07.2020 23:30
DDGLC_Lexicon_of_Greek_Loanwords_in_Coptic-v2-2020.xml	5,35	22.07.2020 23:22
Coptic_Lemma_Schema-v1.2.xsd	0,29	21.07.2020 15:56