

Jerusalem, 2016 February 10

Coptic: A language without words

MARTIN HASPELMATH

Max Planck Institute for the Science of Human History & Leipzig University

Conclusion:

- The “word” notion is not helpful for describing Coptic.
- But Coptic is not necessarily unique – more generally, the “word” notion is not as crucial as is often thought for describing languages.
- We need different descriptive categories for each language, because each language must be described in its own terms.
- Languages are not made out of the same building blocks, unlike things in the world, which are all made out of the same chemical elements.

I. Where I come from

empirical universals of language

- Which nontrivial properties are shared by all languages?
- Which connections can we make between properties of languages?

Nontrivial properties, e.g.

- all languages have special negative morphemes
- all languages have special question constructions
- all languages have demonstratives

Connections between properties, e.g.

- If a language has prepositions rather than postpositions, the possessor follows the noun
- If a language allows one to say “They introduced me to him”, then it also allows “They introduced him to me” (using the same or a shorter construction)

cf. French

Ils me l’ont introduit.

**Ils me lui ont introduit. (Ils m’ont introduit à lui.)*

(Haspelmath 2004)

2. How universal are words?

“All languages have words” (Radford et al. 1999: 145) – a trivial universal?

“Coptic lacks words” – a sensational discovery?

No: Coptic is a fairly usual language, but it can be used well to illustrate the point that linguistics does not necessarily need a “word” notion to describe languages.

Traditionally, the “word” concept is the basis for some other commonly used concepts:

- Syntax describes the combination of *words* to yield sentences.
- Morphology describes the combination of morphemes to yield *words*.
- Syntax and morphology are different “levels” of language description.

But all this presupposes that we know what “words” are. The problem is that we don’t.

3. Attitudes toward words

(1) Words can be identified by speakers, because they have intuitions about words.

Coseriu (1964: 141-142) simply asserts: "Nous estimons la notion de 'mot' comme intuitivement établie" ["We regard the notion of word as intuitively established"]

Aronoff & Fudeman (2005: 36), "speakers – literate and illiterate – have clear intuitions about what is and what isn't a word"

But what if their language does not have a word for ‘word’?

English *word*, French *mot*, Russian *slovo*, Hebrew *mila* – but what is the word for ‘word’ in Coptic?

in Lezgian: *č'al* ‘speech, talk, language, word’

(2) Words can be identified by grammarians, because they all share certain grammatical properties. They form the basis for the distinction between morphology and syntax (*mainstream linguistics*)

(3) Many languages have words, but some don’t (*non-mainstream linguistics*)

(4) Words cannot be properly defined and are irrelevant for linguistics (*my position*)

We certainly know what TYPICAL WORDS are, e.g. ROOTS like *tree, moon, blue, write, laugh*, and TYPICAL COMPLEX WORDS:

verbal roots with tense and person suffixes:

German	<i>kauf-te-st</i>	'you bought'
Latin	<i>scribe-ba-s</i>	'you wrote'
Hebrew	<i>katav-ti</i>	'I wrote'

nominal roots with number and case suffixes:

German	<i>Kind-er-n</i>	'to children'
Latin	<i>puer-is</i>	'to the boys'
Arabic	<i>ḥayawaan-aat-i</i>	'of the animals'

verbal roots with derivational prefixes:

Latin	<i>pro-ducere</i>	'lead forth'
German	<i>ver-kaufen</i>	'sell'

nominal roots with derivational suffixes:

English	<i>kind-ness</i>	
Hebrew	<i>maḥšev-on</i>	'calculator'

But there are also a number of problems in familiar languages, e.g.

English 's genitive:

the men's house, the boys' house, the men who came's house

German infinitival zu:

zu schreiben 'to write', *umzuschreiben* 'to rewrite' (*um-zu-schreiben*)

French preposition de:

la maison de la reine 'the queen's house',
*la maison **du** roi* 'the king's house'

Russian preposition k/ko:

k drugu 'to the friend'
ko mnogim druz'jam 'to many friends'

Such “problems” tend to be found in all languages; they are more noticeable in languages which do not have an established orthography, or in languages whose orthography does not divide a sentence into words, e.g.

在語言學中，單詞是能獨立運用並含有語義內容或語用內容（即具有表面含義或實際含義）的最小單位。

One can suspect that the “problems” arise when a language has complex expressions that are not very much like the typical complex words of Latin or German.

5. Defining and identifying words

So how do mainstream linguists define words?

The answer is: they don't. They generally just assume that words can be defined.

Zedler (1749): "Wort: ein vernemlicher Laut, der etwas bedeutet"

[word: a perceptible sound that means something]

Sapir (1921: 34): "the smallest, completely satisfying bits of isolated 'meaning' into which the sentence resolves itself"

Bloomfield (1933): "a free form which does not consist entirely of (two or more) lesser free forms; in brief, a word is a *minimum free form*"
(free form = a form that can occur on its own a complete utterance)

Wikipedia (2016-02-05):

“In linguistics, a word is the smallest element that may be uttered in isolation with semantic or pragmatic content (with literal or practical meaning).”

But: some compound words are not minimal free forms (e.g. *firewater*: *fire* and *water* are free forms), and many verbs cannot be used on their own, e.g. **put* or **tell*: cf. *put it there, tell me*). Even Bloomfield recognized that the criterion does not work for *the* (*the house*), which is not a free form, but we want to say that it is a word.

Dixon & Aikhenvald (2002: 19-23):

A grammatical word consists of a number of grammatical elements which:

- (a) always occur together, rather than scattered through the clause (the criterion of cohesiveness);
- (b) occur in a fixed order;
- (c) have a conventionalised coherence and meaning.
- (d) Morphological processes involved in the formation of words tend to be non-recursive. That is, one element will not appear twice in a word.
- (e) There will be just one inflectional affix per word.

But: articles, quantifiers, numerals and adjectives occur in a fixed order in many languages:

all these three five little houses (never scattered, always in this fixed order)

6. Complex forms in Coptic

Coptic has what looks like inflectional and derivational prefixes and suffixes, e.g.

derivational:

<i>hôtb</i>	‘kill’	<i>ref-hôtb</i>	‘murderer’
Ⲫⲱⲧⲃ		ⲣⲉϥⲪⲱⲧⲃ	
<i>monak^hos</i>	‘monk’	<i>mnt-monak^hos</i>	‘monasticism, monkhood’
ⲙⲟⲛⲁϭⲟⲥ		ⲙⲛⲧⲙⲟⲛⲁϭⲟⲥ	

inflectional:

1 SG	<i>t^l-sôtm</i>	<i>ne-i-sôtm</i>	ⲧⲢⲱⲧⲙ	ⲚⲈⲒⲢⲱⲧⲙ
2 SG.M	<i>k-sôtm</i>	<i>ne-k-sôtm</i>	ⲕⲢⲱⲧⲙ	ⲚⲈⲕⲕⲱⲧⲙ
3 SG.M	<i>f-sôtm</i>	<i>ne-f-sôtm</i>	ϥⲢⲱⲧⲙ	ⲚⲈϥⲢⲱⲧⲙ
1 PL	<i>tn-sôtm</i>	<i>ne-n-sôtm</i>	ⲧⲚⲢⲱⲧⲙ	ⲚⲈⲚⲢⲱⲧⲙ
2 PL	<i>tetn-sôtm</i>	<i>ne-tetn-sôtm</i>	ⲧⲈⲧⲚⲢⲱⲧⲙ	ⲚⲈⲧⲈⲧⲚⲢⲱⲧⲙ
3 PL	<i>se-sôtm</i>	<i>ne-u-sôtm</i>	ⲢⲈⲢⲱⲧⲙ	ⲚⲈϥⲢⲱⲧⲙ
	‘I hear’ etc.	‘I heard’ etc.		

But there are many more affix-like forms that behave exactly like the tense and person “prefixes” in Coptic, e.g.

– definite articles	<i>p-rôme</i>	‘the man’	<i>te-shime</i>	‘the woman’
	ⲡⲣⲱⲙⲈ		ⲧⲈⲢⲢⲓⲙⲈ	
– demonstratives	<i>pei-rôme</i>	‘this man’	<i>tei-shime</i>	‘this woman’
	ⲡⲈⲒⲣⲱⲙⲈ		ⲧⲈⲒⲢⲢⲓⲙⲈ	
– prepositions	<i>hn-n-rôme</i>	‘in human beings’		
	ⲪⲚⲚⲣⲱⲙⲈ			
	<i>mn-ne-n-šêre</i>	‘with our children’		
	ⲙⲚⲚⲈⲚⲨⲉⲣⲈ			
	<i>e-p-çœeis</i>	‘to the lord’		
	ⲈⲡϨⲞⲈⲒⲢ			

– complementizer	<i>čë-e-tetne-pisteue</i>	‘that you believe’
	ΔΕΕΤΕΤΝΕΠΙΣΤΕΥΕ	
– coordinator	<i>kas hi-sarx</i>	‘bones and flesh’
	ΚΑΣΙΣΑΡΞ	
– relativizer	<i>p-šên et-nanou-f</i>	‘the tree that is good’
	ΠΨΗΝΕΤΝΑΝΟΥϚ	
– existential verb	<i>mnte-pneuma kas hi-sarx</i>	
	lack-spirit	bone and-flesh
	‘Spirits do not have bones and flesh.’	
	ΜΝΤΕΠΝΕΥΜΑΚΑΣΙΣΑΡΞ	

Are these all prefixes? And what about verbs? Are they also prefixes?

<i>neč-ou-noc</i>	<i>n-hroou</i>	‘utter a loud cry’
throw-a-big	ATT-cry	
ΝΕΧΟΥΝΟΒΝΑΡΟΥ		
<i>tse-i-ou-hmč</i>		‘make me drink vinegar’
make.drink-1SG-a-vinegar		
ΤΣΕΙΟΥΣΜΧ		

7. How to describe Coptic

Coptic has several different types of morphemes:

- always stressed morphemes: almost all nouns (e.g. *rôme* ‘man’)
- always unstressed morphemes: many different kinds of grammatical morphemes
- variable morphemes: many transitive verbs, plus a few others (e.g. *nouče/neč-* ‘throw’, *tso/tse-* ‘make s.o. drink’)
- floating morphemes: particles such as *de*

Unstressed morphemes must attach to some stressed morpheme, either at the end, or (much more frequently) at the beginning.

Combinations of stressed morphemes plus attached unstressed morphemes are called **“bound groups”** by Layton (2004) (Haspelmath (2015a: 125) calls them “stress groups”).

Unstressed morphemes may only contain the vowels *e* or *a*, or some syllabic sonorant (liquid, nasal or glide; *y* and *w* are written like *i* and *u*).

The exact sequence and combinability of morphemes is complex, but there is no “word” notion that is helpful for describing Coptic grammar.

Rules for word division in Coptic are arbitrary and differ among authors. Coptic manuscripts do not contain any word divisions. Word divisions are introduced to make Coptic look more like Latin or English.

Should Coptic be written like “isolating languages”, with spaces between all morphemes? e.g.

- b. $\Sigma\text{M } \Pi \text{ ME}\Sigma \text{ } \Psi\text{OMNT}$
hm-p-meh-šomnt [həm-p-məh-ʃómənt]
 in-DEF.M-ORD-three
 ‘in the third’
- c. $\text{OY } \rho\epsilon\epsilon\text{q } \rho \text{ NOBE}$
ou-ref-r-nobe [w-rəf-ər-nóbə]
 INDF-AGT-do-sin
 ‘a sinner’
- d. $\epsilon \text{ } \Upsilon\text{NT } \text{C } \text{OY } \Psi\epsilon\epsilon\rho\epsilon$
e-unt-s-ou-šeeere [ə-wənt-əs-w-ʃéʔrə]
 REL-have-3SGF-INDF-daughter
 ‘who had a daughter’
- e. $\text{MN } \text{N } \epsilon\text{TE } \text{N } \text{TN } \text{N}\lambda\gamma \text{ } \epsilon\rho\text{O } \text{OY}$
mn-n-ete-n-tn-nau *ero-ou* [mən-n-ətə-n-tən-náw ərów]
 with-DEF.PL-REL-NEG-1PL-see to-3PL
 ‘and those which we do not see’

For linguists, it does not matter how Coptic is written. Probably for our reading habits, it is easiest if it is written with hyphens linking unstressed and stressed morphemes (much as in Layton’s grammar).

Crucially, we cannot equate stress groups with words, because stress groups often contain elements that have NO MORPHOSYNTACTIC COHERENCE, e.g.

- (52) a. $\alpha\text{PA}\gamma\text{ΛOC } \text{N}\lambda\gamma \text{ } \epsilon\rho\text{OY}$
a-paulos nau ero-f [a-páwlos náw əró-f]
 PRET-Paul see to-3SGM
 ‘Paul saw him.’
- b. $\text{ΠENT}\alpha\text{M}\omega\gamma\text{CHC } \text{C}\Sigma\alpha\text{I } \epsilon\text{TBHHT}\rho$
p-ent-a-môusês shai etbêêt-f [p-ənt-a-mo:isé:s sháj ətbé:ʔt-əf]
 DEF.M-REL-PRET-Moses write about-3SGM
 ‘the one of whom Moses wrote’

And verbs can sometimes be unstressed:

- (51) a. ΠΕΧΕΙΗCOYC
pečē-iēsous [pəkʰə-je:sú:s]
 say-Jesus
 'Jesus said.'
- b. ΕΝΕΧΤΕQCZIME ΕΒΟΛ
e-neč-t-ef-shime *ebol* [ə-nəkʰ-t-əf-shí:mə əból]
 to-throw-DEF.F-3SGM-wife out
 'to throw out (=divorce) his wife'

8. Other authors on Coptic

“structuralist authors“:

Shisha-Halevy (2002: 429-430):

“Word division, entirely strange to Coptic and Egyptian analytic sensibility, accommodates the western student’s bias, in a subjective and imprecise conception, suited to the (Indo-) European typical parcel unit of lexical stem and grammemic affixes, packaged as a ‚word‘, which often conflicts with the Coptic native typology of lexical units preceded by affixes.... The word is no longer a notion of general analytical significance.“

Layton (2004: §27):

“Coptic morphs – the building blocks of patterns and constructions – typically occur in strings, which are united by relationships of adjacent dependency or ‚boundness‘. Such morph strings will be called BOUND GROUPS. ...The constituents, sequence, and boundaries of a bound group are regulated by the basic dependency properties of each morph that happens to occur within it, as well as by other factors...”

generative and typological authors:

Reintges (2001: 177):

“I will argue that the person, number and gender markings on Coptic verb forms are not agreement affixes, but pronominal clitics.”

Egedi (2007: 118)

“In view of the above demonstrated arguments, my suggestion is to avoid the traditional use of such category terms as suffix, infinitive and inflection or conjugation in the Coptic language. It would be more adequate to utilise such expressions as weak pronouns or clitic pronouns, absolute, pronominal, etc. forms of the verb without referring to any kind of finiteness, instead.”

Grossman & Polis (2015: 8):

“Coptic has prefixed subject affixes on verbs. While they may follow TAM/Polarity affixes, they always precede the lexical verb.”

a-f-čī *n-ou-oik*
 PST-3SG.M-takeACC-INDEF-bread
 ‘He took some bread.’

But cf. *a-p-rôme* *čī ...* (cf. their criteria: uninterruptability, adjacency)
 PST-DEF-man take
 ‘The man took...’

9. Combining the criteria: Persuasion by test batteries

- often in linguistics, a number of criteria are selected and applied, and in the published accounts usually all of them point in the same direction
- the more criteria converge, the more persuasive the argument becomes (Haspelmath 2011: §4)

	Zwicky & Pullum 1983	Kanerva 1987	Bresnan & Mchombo 1995	Ackerman & LeSourd 1997	Monachesi 1999	Harris 2000	Milićević 2005	Lieber & Scalise 2006	Bickel et al. 2007
Free occurrence				x			x		
External mobility and internal fixedness	x			x	x	x			
Uninterruptibility				x					x
Non-selectivity	x	x			x	x	x		x
Non-coordinatability			x	x	x		x	x	x
Anaphoric islandhood			x					x	
Nonextractability			x					x	
Morphophonological idiosyncrasies	x	x			x	x	x		
Deviations from biuniqueness									x

Table 1. *Nine studies that examine wordhood using test batteries*

But since different criteria are applied in different languages, the procedure is non-rigorous.

10. Diagnostics vs. criteria

Is this not a serious problem?

Zwicky (1985: 284-286): no, because diagnostics are only hints, they are not defining criteria

diagnostics	criteria
(= symptoms)	(= definientia)
– point us to an underlying reality that may not be observable directly	– jointly necessary and sufficient
– may be absent or occasionally contradict each other	– when they are applicable, they can be used to classify the phenomena
– ultimately, taxonomy (= theory-building) is done on the basis of the researcher's intuitions	– when the application is unclear, sharpen the criteria or get more data

“analogous to medical diagnosis; as in medical diagnosis, interfering factors can prevent even clear cases from exhibiting a certain symptom, and a particular symptom might result from some condition other than the one at issue” (Zwicky 1985: 284-286)

But: in medicine, there is abundant evidence that “every disease should NOT be described in its own terms” – many or most diseases are caused by a factor that is independent of the symptoms (pathogens), so an apriorist approach is sensible.

In language, the postulation of cross-linguistic (universally applicable) categories is not so much based on evidence, as on hope: That we will eventually be able to identify the correct categories of UG.

If intuitions differ across researchers, there is no way to resolve the disagreements.

1.1. Some positive steps: Comparative concepts

To compare languages rigorously, we need COMPARATIVE CONCEPTS – a special set of concepts that are intended for comparison, not description/analysis (Haspelmath 2010).

plenimorph = a morph that denotes a thing, a process or a property (= a root)

minimorph = a morph with a meaning that is normally omitted in translation into some other language without significant loss of content (i.e., a meaning that could easily be inferred from context, or a meaning that makes a small, subtle contribution)

Universal I: In all languages, plenimorphs are longer on average than minimorphs. (In fact, almost all languages have many plenimorphs that are bisyllabic or longer, and all have many minimorphs that are monosyllabic or shorter.)

Universal II: In all languages, plenimorphs show greater ordering variability than minimorphs. (In fact, all languages have some minimorphs whose ordering is strictly fixed with respect to a related plenimorph.)

Universal III: In all languages, the coalescence properties of minimorphs (prosodic dependency, adjacency, narrow scope, shape idiosyncrasy) correlate strongly with each other: If a morph is more coalescent than another morph with respect to one of the properties, it also tends to be more coalescent with respect to the other properties.

These concepts are not necessary for describing Egyptian or Coptic. They are not intended as an aid to language description.

Specialists of particular languages should certainly be inspired by typological work, but should not feel the need to adopt all their concepts from the typologists. Typology and descriptive linguistics should be in symbiosis, but not in a one-sided dependency relation.

References

- Aronoff, Mark & Fudeman, Kirsten. 2005. *What is morphology?* Oxford: Blackwell.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt.
- Coseriu, Eugenio. 1964. Pour une sémantique diachronique structurale. *Travaux de linguistique et de littérature* 2.1: 139-186.
- Dixon, R.M.W. & Aikhenvald, Alexandra Y. 2002. Word: a typological framework. In: Dixon, R.M.W. & Aikhenvald, Alexandra Y. (eds.) *Word: a cross-linguistic typology*. Cambridge: Cambridge University Press, 1-41.
- Egedi, Barbara. 2007. Reconsidering the categorial status of the Coptic suffix and conjugation base. *Proceedings of the Fourth Central European Workshop of Young Egyptologists*.
- Grossman, Eitan & Polis, Stéphane. 2015. Dispreferred structures through language change: the diachrony of affix ordering. Handout, *Workshop on Verbs, Verb Phrases & Verbal Categories*, HUJI 23-24 March 2015
- Haspelmath, Martin. 2004. Explaining the Ditransitive Person-Role Constraint: A usage-based approach. *Constructions* 2. (online)

- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80.
- Haspelmath, Martin. 2015a. A grammatical overview of Egyptian and Coptic. In Eitan Grossman, Martin Haspelmath & Tonio Sebastian Richter (eds.), *Egyptian-Coptic linguistics in typological perspective*, 103–143. Berlin: De Gruyter Mouton.
- Haspelmath, Martin. 2015b. Defining vs. diagnosing linguistic categories: A case study of clitic phenomena. In Joanna Błaszczak, Dorota Klimek-Jankowska & Krzysztof Migdalski (eds.), *How categorical are categories? New approaches to the old questions of noun, verb and adjective*. Berlin: De Gruyter Mouton.
- Layton, Bentley. 2004. *A Coptic grammar: With a chrestomathy and glossary: Sahidic dialect*. 2nd edition. Wiesbaden: Harrassowitz.
- Radford, Andrew, Martin Atkinson, David Britain, Harald Clahsen & Andrew Spencer. 1999. *Linguistics: an introduction*. Cambridge: Cambridge University Press.
- Reintges, Christ. 2001. Aspects of the morphosyntax of subjects and objects in Coptic Egyptian. *Linguistics in the Netherlands 2001*, 177-188.
- Sapir, Edward. 1921. *Language: an introduction to the study of speech*. New York: Harcourt Brace.
- Shisha-Halevy, Ariel. 2002. A definitive Sahidic Coptic grammar: Review article of Layton (2000). *Orientalia Nova Series* 71(4). 424-459.
- Zedler, Johann Heinrich (ed.) 1749. *Grosses vollständiges Universal-Lexicon Aller Wissenschaften und Künste*, vol. 59 (Wor–Wuq). Halle und Leipzig: Johann Heinrich Zedler.
- Zwicky, Arnold M. 1985. Clitics and particles. *Language* 61(2). 283–305.