Wordhood domains in Central Kurdish

Shuan Osman Karim
University of Cambridge
sk2300@cam.ac.uk

Erika Just
University of Zurich
erika.just@uzh.ch

Follow Along

Presentation in a nutshell

- Word is not a cross-linguistically consistent concept
- Wordhood criteria vary according to: no internal pauses, language-specific p-word requirements, free occurrence, non-interruptibility, positional freedom, non-selectivity, extractability, anaphoric islandhood
- Often there are multiple cohesion domains within a single language
- Wordhood progress in typology over last few decades
- A possible way to go about it: multivariate approach
- (Suleymani) Soranî has several such domains in terms of p-word, g-word.
- I add to this the o-word which reflects a combination of native-speaker intuitions and prescriptive conventions.

What is word?

- No fruitful attempt at establishing “word” as a comparative concept (Haspelmath 2011)
- A word is (i) a free morph, or (ii) a clitic, or (iii) a root or a compound possibly augmented by non required affixes and augmented by required affixes if there are any (Haspelmath 2023: 294).
- “this is not a natural definition … there is a widespread implicit view in linguistics that the unit word is a natural way of dividing up texts in languages.”
- “However, it is quite possible that the salient general notion of word is an artifact of European spelling conventions”
- “it could be that the notions of ‘word’ ,‘morphology’ and ‘syntax’ are stereotypes derived from our tradition that have no actual basis in the reality of languages.”

What is word?

- Diagnostics have to be fine-grained and logically independent (van Gijn & Zúñiga 2014, Bickel & Zúñiga 2017, Tallman 2019)
- Significant advancements in recent years in understanding this concept
- Definition has been effectively broken down into phonological and morphosyntactic components (going back to Dixon 1977)
- To which we add consideration of the orthographic component
- Instead: being agnostic of the possibility of different wordhood-domains The degree of convergence is an empirical question

Domains in Kurdish

- previous accounts of Kurdish already point towards wordhood wrt the verb seems to be gradient
- applicatives have been called “absolute prepositions” because of their (sometimes) prepositional etyma (Karim and Salehi 2022)
- especially preverbs are treated inconsistently: often treated as syntactically independent structures despite their dependence on the verbal stem
- CPMs are called clitic person markers despite evidence pointing to their being more affix-like (Wackernagel Affixes? Nevis and Joseph 1993)

No isomorphism on different levels

(p-word: stress-PW vowel harmony-PW)

- However, even within different levels of analysis, languages might differentiate between various non-isomorphic subcategories
- True for morphosyntactic criteria, but also for phonological criteria
  e.g. stress vs. vowel harmony in Turkish (Kabak & Vogel 2001)

  [[*sev-il*]-*mi*]-*yor-uz*
  love-PASS-NEG-PROG-1PL
  'We are not loved.'

The g-word

“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.”

(Dixon & Aikhenvald 2002: 19)

The g-word

"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."

(Dixon & Aikhenvald 2002: 19)

Soranî

a. *pirsyar=yan* ***lê=****kird-în*
question=3pl.A **ABL.AP=**LV.PST-1PL.P
'They asked us a question.'

b. ***lê=****yan* *da-∅*
ABL.AP=3PL.A give.PST-3SG.P
'they set out (lit. hit it)'

The g-word

"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."

(Dixon & Aikhenvald 2002: 19)

c. *dît-mân-in*
see-1PL.**A**-3PL.**P**
'we saw them'

d. *dît-in-î*
see-3PL.**P**-3SG.**A**
's/he saw them'

e. *da-î(-∅)-n-ê*
give-3SG.**A**(-3SG.P)-3PL.**O**$_{AP}$-DAT.AP
's/he gave it to them'

The g-word

"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."

(Dixon & Aikhenvald 2002: 19)

"while the meaning of a word is related to the meanings of its parts, it is often not exactly inferable from them." E.g., blackbird vs. black bird.

gułî sur vs. gułe sur

"the listedness of at least SOME such units is probably viewed by many as a necessary condition to establish the members of the class of such units as words (Harris 2000: 599).

hełkewtin vs. *serkewtin*

A multivariate approach to word domains

Possible criteria (cf. Bickel & Zúñiga 2017)

- Items in a domain are characterised by
  - inflection:
    - items are required for the verb to be appropriately used as an independent syntagma in a sentence
    - dedicated, productive and obligatory use in certain settings
  - selection:
    - items cannot occur on their own and require a verb as a (potential) host
    - only captures the specific combinatorial requirements of an element: distinct from independent pronounceability and from the syntactic behavior in terms of adjacency, linear order, or position

A multivariate approach to verb domains

Possible criteria (cf. Bickel & Zúñiga 2017)

- Morphosyntactic cohesion: how items stick together (or not)
  - promiscuity wrt the host: can there be another besides the verb?
  - cross-slot dependency: do individual items within a domain depend on each other?
  - insertion potential: can other elements be interspersed? (Exception: they select the verb or the items, or are required by the inflection)
  - fixed arbitrary position: is the relative order of the item and the stem is fixed arbitrarily (as opposed to be driven by semantics or information structure)?

Morphosyntactic cohesion in the Soranî verb

- **promiscuity wrt the host:**
- cross-slot dependency
- insertion potential
- Fixed arbitrary position

f. nan-ekan**=im** xward
bread-DEF.PL=1SG.A eat.PST

g. xward**-im**-in
eat.PST-1SG.A-3PL.P

h. ktêb-eke=m **lê**=t wer-girt
book-DEF.SG=1SG.A from=2SG PV-take.PST

i. ktêb-eke=m **lê**-wer-girt-î
book-DEF.SG=1SG.A ABL.AP-PV-take.PST-sSG.APO

Morphosyntactic cohesion in the Soranî verb

- promiscuity wrt the host
- **cross-slot dependency**
- insertion potential
- fixed arbitrary position

f.	**nan-ekan=im**	xward**-im-in**
	bread-DEF.PL=1SG.A	eat.PST**-1SG.A-3PL.P**
g.	**nan-ekan=im**	xward**-im-in**
	bread-DEF.PL=1SG.A	eat.PST**-1SG.A-3PL.P**

‘I ate the bread/it’

h.	ktêb-eke=m	**lê=t**	wer-girt
	book-DEF.SG=1SG.A	**from=2SG**	PV-take.PST
i.	ktêb-eke=m	**lê-**wer-girt**-î**	
	book-DEF.SG=1SG.A	**ABL.AP-**PV-take.PST**-2SG.APO**	

‘I took the book from you’

Morphosyntactic cohesion in the Soranî verb

- promiscuity wrt the host
- cross-slot dependency
- **insertion potential**
- fixed arbitrary position

(✔) nan-ekan(❌)=im (✔) xward (✔)

(✔) xward(❌)-im(❌)-in (✔)

(✔) ktêb-eke(❌)=m (✔) lê(❌)=t (✔)wer-(❌)girt (✔)

(✔) ktêb-eke(❌)=m (✔) lê-(❌)wer-(❌)girt(❌)-î (✔)

(dwênê) ‘yesterday’

Morphosyntactic cohesion in the Soranî verb

- fixed arbitrary position $\alpha \rightarrow \beta$ $\alpha \rightarrow \beta$ $\alpha \rightarrow \neg(\beta)$

(NP)	[(EM)]	**[(CPM)]/ [(CPM)]**	**(AP)**	(PV)	(NEG)	(TAM)	$	(PAS)		PN			(DIR)/ (ITR)/ **(AP)**
(NP)	[(EM)]	[CPM]	**(AP)**	(PV)	(NEG)	(TAM)	$	(PAS)		**(PN)**	**(PN)**		(DIR)/ (ITR)/ **(AP)**
(NP)	[(EM)]	[CPM]	**(AP)**	(PV)	(NEG)	(TAM)	$	(PAS)	TAM	**(PN)**	**(PN)**		(DIR)/ (ITR)/ **(AP)**
(NP)	[(EM)]	[CPM]	**(AP)**	(PV)	(NEG)	**TAM**	$	(PAS)		**(PN)**	**(PN)**	**COND**	(DIR)/ (ITR)/ **(AP)**

Phonological cohesion domains in the Soranî verb

caveat: in-depth work on phonological/prosodic cohesion in Kurdish still needs to be done, but we do know about some basics

- Stress Placement following McCarus 1997, 2009:

  Verbal: NEG > Preverb > SBJ/IMP > Stem(-final V)

  Nominal: DEF > Stem(-final V)

- Stress Placement Revised:

  [ˈ**Ktê**bim lê wer girtî.] — [ˈ**Ktê**bim wer girt] [ˈ**lêt**.] — [ˈ**Ktê**bim [ˈ**lêt**] wer girt.]
  [Ktêbim lê wer ˈ**ne**girtî.] — [Ktêbim wer ˈ**ne**girt] [ˈ**lêt**.] — [Ktêbim [ˈ**lêt**] wer ˈ**ne**girt.]

Speaker intuition: orthography

- As many Soranî speakers received formal education in other languages, they show idiosyncratic orthographic variation.
- Hypothesis: Soranî orthography reflects speaker intuition about wordhood.
- Compare: لێمەرگرتی* – لێم وەرگرتی – لێم وەر گرتی
- Compare: وەرمگرت لە تۆ* – وەرم گرت لە تۆ
- Compare: ئەیشزانی – ئەیش زانی
- Compare: پێم پێ ئەکەنن – پێم پێئەکەنن* – پێم پێ کەنین – پێمپێکەنین* – پێم پێکەنین

Discussion

- Need for more fine-grained language-specific studies for a bigger picture:
  - Are there cross-linguistic tendencies for particular domains to nest?
  - Are there differences in terms of stability for different domains?
    E.g. there is evidence from Welsh suggests that phonological (verbal) wordhood-domains are more stable than morphosyntactic ones (Dedio to appear)
  - What role do different domains play in language planning, processing and L1 acquisition
    E.g. it is often argued that children under-segment the input, and extract units “larger than words” (Arnon 2021)

References

Bauer, Laurie. 1988. I*ntroducing linguistic morphology*. Edinburgh: Edinburgh University Press.

Bickel, Balthasar & Fernando Zúñiga. 2017. The 'word' in polysynthetic languages: Phonological and syntactic challenges. In Michael Fortescue, Marianne Mithun and Nicholas Evans (eds.), *The Oxford handbook of polysynthesis*, 158-186. Oxford: Oxford University Press.

Bickel, Balthasar, Kristine A. Hildebrandt & and René Schiering. 2009. The distribution of phonological word domains: a probabilistic typology. In Janet Grijzenhout & Baris Kabak (eds.) Phonological Domains: Universals and Deviations, 47-78. Berlin, New York: De Gruyter Mouton.

Dixon, Robert M.W. 1977. Some Phonological Rules in Yidin[y]. *Linguistic Inquiry* 8 (1), 1-34.

Harris, A. C. 2000. Where in the Word is the Udi Clitic? Language, 76(3), 593–616. https://doi.org/10.2307/417136

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia linguistica* 51, 31-80.

Haspelmath, Martin. 2023. Defining the word, WORD, 69:3, 283-297, DOI: 10.1080/00437956.2023.2237272

Karim, Shuan Osman and Salehi, Ali. (2022,) An applicative analysis of Soranî "absolute prepositions". In Applicative Morphology: Neglected Syntactic and Non-syntactic Functions. Sara Pacchiarotti and Fernando Zúñiga eds., Berlin, Boston: De Gruyter Mouton, 263-298. https://doi.org/10.1515/9783110777949-010

McCarus, Ernest N. 1997. Kurdish Phonology. In Kaye, Alan S. (ed.), no booktitle, 691-705. Winona Lake: Eisenbrauns.

Mugdan, Joachim. 1994. Morphological units. In R. E. Asher (ed), *The encyclopedia of language and linguistics,* 2543–2553. Oxford: Pergamon Press.

Nevis, J.A., Joseph, B.D. (1993). Wackernagel affixes: evidence from Balto-Slavic. In: Booij, G., van Marle, J. (eds) Yearbook of Morphology 1992. Yearbook of Morphology. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-3710-4_4

Tallman, Adam J.R. 2020. Beyond grammatical and phonological words. *Language and Linguistics Compass* 14(2).

van Gijn, Erik & Fernando Zúñiga. 2014. Word in the Americanist perspective. *Morphology* 24, 135–160.