

A morphological analyzer for Wolof using finite-state techniques

Bamba Dione, University of Bergen
Ines Fiedler, Humboldt-Universität zu Berlin

0. Introductory remarks

- cooperation between project D1 of the SFB 632 (Amir Zeldes), Bamba Dione (Wolof mother tongue speaker and computer linguist), and our project
- it fits in our plan to analyze how predicate-centered focus types are used in spontaneous natural texts in selected sample languages for which sufficiently large corpora are available – one of these languages is Wolof, another one is Hausa, for which a corpus was developed within the SFB by projects B1, A5 and D1
- methodological assumption: predicate-centered focus is largely absent from narrative texts and occurs there largely in direct speech, while dialogic discourse provides more relevant tokens → it is not an easy task to develop such a corpus on the basis of written material on which one can undertake semi-automatic search for special structures
- we try that by building a corpus on the basis of the Wikipedia Wolof; we decided for the Wikipedia-Wolof corpus for two reasons:
 - sufficient quantity of written material (open source)
 - sufficient quality of texts
- but: only one text type, even though from different domains and different authors – advantage compared to the bible translation used by Bamba before
- as focus structures are strongly grammaticalized in Wolof, we expect nevertheless to get enough material out of the texts to come to a final conclusion in which contexts the so-called “Verb emphatic” verb form (Robert 2010) in Wolof texts occurs
- in the following, some general information on the language is given, with focus on the morphological forms which might be relevant for parsing

1. Wolof¹

+ general information

- ISO 639-3: wol
- Niger-Congo, Atlantic, Northern, Senegambian, Fula-Wolof, Wolof

¹ The following information is mainly taken from Fiedler & Terwite 2011 (Ms.)

- spoken chiefly in Senegal (about 3.9 mio speakers) with a small number (approx. 100,000) in Mauretania. Closely related to Gambian Wolof (ISO 639-3: wof); the dialects are mutually understandable and “differences are chiefly in the domain of vocabulary and phonology.” (Dunigan 1994: 7)
- Wolof functions as the lingua franca of Senegal, with 80% speaking it (sources vary on the actual number of proficient speakers (L1 + L2), which according to some (cf. Mc Laughlin 2009: 1184) may be as high as 7 million and “quite possibly more”.
- has the status of an official language in Senegal

+ typological information

- morphological type: agglutinative
- word order type: SVO
- no lexical tone, no tonal accent, no pitch accent, stress does not correlate with pitch, intonation not used for focalization (Rialland & Robert 2001)
- vowel harmony system (Advanced Tongue Root feature (ATR)) which is reflected in the orthography (see ex. (2) and (3)) and may create problems for parsing

(1)	[+ATR]		[-ATR]	
[i]	- <i>	[u]	- <u>	
[e]	- <é>	[o]	- <ó>	
		[ɛ]	- <e>	
		[ə]	- <ë>	
			[a]	- <à>
			[ʌ]*	- <a>*

* [ʌ] is not given in all descriptions of Wolof (cf. Mc Laughlin 2009, Russell 2006)

(2)	lekk-oon-ngeen	[ɔŋŋɛɛn]	-ATR
	eat-PST-PERF.2P		
	y'all ate		
(3)	dóór-óón-ngéén	[oonŋgeen]	+ATR
	hit-PST-PERF.2P		
	y'all hit		

+ Nominal morphology (noun class system)

- noun classes: 8 singular, 2 plural (dialectal variant: in Dakar Wolof reduced system with only one singular and one plural class)
- class markers agree with the nominal stem, demonstratives and the ‘annexion’ marker; bare noun without class marker gets indefinite interpretation; class markers express definite referentiality (here glossed as DET – determiner)

- in Standard Wolof Orthography, they are written separately

Singular			Plural		
xaj	bi	'the dog'	ja	yi	'the markets'
gaal	gi	'the boat'	góór	ñi	'the men'
ndap	li	'the pot'			
wax	ji	'the talk'			
jën	wi	'the fish'			
ndaw	si	'the young woman'			
saw	mi	'the urine'			
nit	ki	'the person'			

(cf. Torrence 2005: 21)

- Wolof has three types of "pronouns":

- a: subject agreement marking pro-forms which go with the verb
- b: pronouns functioning as objects
- c: pronouns functioning as subjects (variously called independent pronouns)

Verb paradigms

+ the Wolof verb constituent has two components: an invariant lexical stem and an inflectional marker conveying the grammatical specifications of the verb (person, number, tense/aspect, mood) as well as the information structure of the sentence (focus) (Robert 2010: 237)

- the inflectional marker is preposed, postposed, or suffixed to the lexical stem

Indicative			Subjunctive
Affirmative	Affirmative "emphatic"	Negative	Affirmative
Perfect (PERF) verb na-pro	Subject Emphatic (FOC_1) SUBJ (pro)-a verb	Negative (NEG) verb- u-pro	Obligative na -verb
Presentative (PRES) pro-ng-i verb	Complement Emphatic (FOC_2) COMPL l-a-(pro) verb	--	Imperative verb -(a)l
Aorist (AOR) (pro)-ø verb	Verb Emphatic (FOC_3) da-pro verb	Emphatic Negative (NEGFOC) d-u-pro verb	

Table 3: Formal characteristics of Wolof verb paradigms (Robert 2000, 2010)

- auxiliaries in Wolof constitute an amalgamated form, consisting of different morphemes:

- presentative: ng + i; ng = ?; i = spatial determination of verb
 - subject emphatic: -a = identificational marker
 - complement emphatic: l-a-pro; l = expletivum, a = identificational marker
 - verb emphatic: da-pro; da = di = equative copula
 - negative emphatic: d-u-pro; d = di = equative copula; u = negation marker
- (cf. Kihm 1999: 247 for a discussion of the different copulae; see also Torrence Ms.)

'Perfect'

- (4) xale yi lekk **na-ñu** ceeb bi
 child DET.P eat PERF-3P rice DET
 The children have eaten the rice. (Zribi-Hertz & Diagne 2002: 829 (6b))
- (5) mburu mi | lekk **na** ko
 bread DET (pause) eat PERF.3S 3S.OBJ
 That bread, he ate it.

'Presentative'

- (6) Peer | mu-**ngi** lekk
 PN (pause) 3S-PRES eat
 (As for) Peer, he is eating.

'Subject Emphatic' → focus on subject

- (7) Daba **moo** ma ko bind.
 PN 3S.FOC_1 1S.OBJ 3S.OBJ write
 {Who wrote you this letter?} DABA wrote it to me.
 (Robert 2010: 254, Robert 2000: 238)

'Subject Emphatic' → entity-centralthetic utterance

- (8) Musaa, moo dóor Ndey
 PN 3S.FOC_1 beat PN
 {What is going on here?} It is Musa who has trashed Ndey.
 (Robert 2010: 254, Robert 2000: 238)

‘Complement Emphatic’ → focus on object

- (9) Peer | mburu mi la lekk
PN (pause) bread DET FOC_2.3S eat
As for Peer, it was bread he ate.

‘Negative’ ; ‘Verb Emphatic’ → focus on semantic content of verb

- (10) tóx-u-ma, da-ma-y fo
smoke-NEG-1S FOC_3-1S-IPFV play
I am not smoking, I am PLAYING (with the cigarette). (Robert 2000 : 259)

‘Verb Emphatic’ → focus on truth value of the statement, confirmation (Robert 2010: 251)

- (11) Da-ma lekk
FOC_3-1S eat
In FACT, I ate ~ it is because I ate that....

‘Verb Emphatic’ in first sentence of a story → event-centralthetic utterance (Robert 1996: 155?)

- (12) Ca jamano yu yàgg, da-fa am-oon
in time REL last_long, FOC_3-3S have/exist-PST
benn ilimaanu jàkka
one imam mosque
(En des temps lointains, il y avait un imam de mosquée. {Il convoqua une palabre à la mosquée. ...})
In former times, there was an imam of a neighborhood mosque.

Problems with these descriptions:

- focus in Wolof is mainly expressed by means of the different conjugation paradigms (cf. Robert 2000, 2010); these are used to express different scopes of focus; the communicative point of this means is mainly contrastive, but also assertive
→ but the name of the inflectional forms does not always correspond to the expressed focus type
- ‘verb emphatic’ expresses SoA focus, but also truth value focus - but most often, it is not used to indicate focus on the verb – it occurs mainly in clause chains and delivers explications for the preceding proposition – like this, it expresses rather wide focus on the whole sentence

- ‘subject emphatic’ is the only means to express focus on the grammatical subject, but is also used to indicatethetic utterances and truth value focus

- ‘complement emphatic’ indicates focus on any element except the subject or the inflected verb; if the verb is nominalized and fronted, then this form is also used
- there is a strong interaction between the kind of predicate and the conjugation paradigm used which leads to an asymmetric distribution:

- stative verbs normally occur together with ‘verb emphatic’, lying assertive focus on the verb, thereby expressing a quality of the subject; with the perfect, stative verbs express a temporary property of the subject – gets a contrastive focus reading
- with action verbs, the ‘verb emphatic’ form may express assertive and contrastive focus

→ we therefore hope to get answers with the help of a large annotated Wolof corpus

Abbreviations

DET	Determiner, singular	P	Plural
DET.P	Determiner, plural	PERF	Perfect
FOC_1	“Subject” emphatic form	PFV	Perfective
FOC_2	“Complement” emphatic form	PN	Proper name
FOC_3	“Verb” emphatic form	PRES	Presentative
IPFV	Imperfective	PST	Past
NEG	Negative	REL	Relativ
OBJ	(Direct) object	S	Singular

References

- Diagne, Pathé. 1971. *Grammaire de wolof moderne*. Paris: Présence Africaine.
- Dunigan, Melynda B.. 1994. *The clausal structure of Wolof: A study of focus and cliticization*. (PhD Thesis). Ann Arbor: UMI.
- Fal, Arame, Rosine Santos & Jean Léonce Doneux. 1990. *Dictionnaire wolof-français. Suivi d'un index français-wolof*. Paris: Karthala.
- Fiedler, Ines & Christoph Terwitte. 2011. Predicate-centered focus in Wolof. Paper presented at a project meeting, Berlin 14.6.2011, Ms.
- Ka, Omar. 1994. *Wolof phonology and morphology*. Lanham/New York/London: University Press of America.
- Kihm, Alain. 1999. Focus in Wolof: A study of what morphology may do to syntax. In *The grammar of focus*, eds. Georges Rebuschi & Laurice Tuller, 245-274. Amsterdam, Philadelphia: Benjamins.
- Mc Laughlin, Fiona. 2009. Wolof. In *Concise encyclopedia of languages of the world*, eds. Edward K. Brown & Sarah Ogilvie, 1184-86. Amsterdam [u.a.]: Elsevier.
- Munro, Pamela & Dieynaba Gaye. 1991. *Ay baati Wolof: A Wolof dictionary* (UCLA Occasional Papers in Linguistics, vol. 9). Los Angeles: University of California, Los Angeles.
- Ngom, Fallou. 2003. *Wolof* (Languages of the world materials 333). München/Newcastle: LINCOM Europa.
- Rialland, Annie & Robert, Stéphane. 2001. The intonational system of Wolof. *Linguistics* 39:893-939.
- Robert, Stéphane. 1991. *Approche énonciative du système verbal le cas du Wolof*. Sciences du langage. Paris: Ed. du Centre National de la Recherche Scientifique.
- Robert, Stéphane. 1996. Aspect zéro et dépendance situationnelle: L'exemple du wolof. In *Dépendance et*

- intégration syntaxique: Subordination, coordination, connexion*, ed. Claude Muller, 153-161. Tübingen: Niemeyer.
- Robert, Stéphane. 2000. Le verbe wolof ou la grammaticalisation du focus. In *Topicalisation et focalisation dans les langues africaines*, ed. Bernard Caron, 229-268. Louvain/Paris: Peeters.
- Robert, Stéphane. 2010. Focus in Atlantic Languages. In *The expression of information structure: A documentation of its diversity across Africa*, eds. Ines Fiedler and Anne Schwarz, 233-260. Berlin: John Benjamins.
- Russell, Margaret A. 2006. The syntax and placement of Wolof clitics. (unpublished PhD thesis). Urbana: University of Illinois at Urbana-Champaign.
- Torrence, Harold. 2003. Verb Movement in Wolof. *UCLA Working Papers in Linguistics* 9:31.
- Torrence, Harold. 2005. On the Distribution of Complementizers in Wolof. (PhD thesis). Los Angeles: University of California at Los Angeles (UCLA).
- Torrence, Harold. n.a. The Morphosyntax of Wolof Clefts: Structure and Movement. Ms.
- Zribi-Hertz, Anne & Lamine Diagne. 2002. Clitic placement after syntax: Evidence from Wolof person and locative markers. *Natural Language & Linguistic Theory* 20, 823-84.