

Prosodic segmentation and grammatical analysis in cross-linguistic corpora

CORPAFROAS



A. Mettouchi & M. Vanhove
The CorpAfroAs team
The CorTypo team



amina.mettouchi@cnrs.fr ; martine.vanhove@cnrs.fr



École Pratique
des Hautes Études

Langage, Langues et
LLACAN
Cultures d'Afrique Noire

- <http://corpafroas.huma-num.fr/>
- <https://www.benjamins.com/#catalog/books/scl.68/main>



John Benjamins Publishing Company

New ▾ Catalog ▾ Online Resources Customer Services ▾ For Authors ▾ JB Antiquariat ▾

You are here: Catalog ▾ Books ▾ SCL ▾ SCL 68

Main Table of Contents Subjects

Corpus-based Studies of Lesser-described Languages

The CorpAfroAs corpus of spoken AfroAsiatic languages

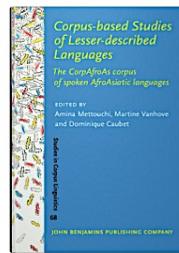
Edited by *Amina Mettouchi, Martine Vanhove and Dominique Caubet*
EPHE (LLACAN), Paris / CNRS (LLACAN), Paris / INALCO (LaCNAD), Paris

This volume presents new findings based on the analysis of spoken corpora in thirteen different Afro-Asiatic languages - a unique endeavor in the domain of lesser-described languages. It will be of interest to corpus linguists, general linguists, typologists, and linguists specializing in Afro-Asiatic languages. In addition to the rarity of corpus studies based on endangered and lesser-described languages, the volume is remarkable due to its focus on the role of prosody in interaction with several other phenomena, including code-switching and borrowing. Phonology, syntax, and information structure are explored, and the issue of the elaboration of strategies for the typological comparison of corpora is addressed in several papers. The volume also contains a presentation of software development conducted within the scope of the CorpAfroAs project and based upon the widely used ELAN. The sound-indexed, and morphosyntactically-annotated corpora, with their OLAC metadata and several other deliverables can be accessed and searched at <http://dx.doi.org/10.1075/scl.68.website>.

[Studies in Corpus Linguistics, 68] Expected May 2015. vi, 338 pp.

Publishing status: Printing
© John Benjamins

DOI: 10.1075/scl.68



Hardbound - Forthcoming 
ISBN 9789027203762 | EUR 99.00 | USD 149.00

e-Book - Forthcoming **Ordering information**
ISBN 9789027268891 | EUR 99.00 | USD 149.00

Table of Contents

Preface

Amina Mettouchi, Martine Vanhove and Dominique Caubet

Part 1: Phonetics, phonology and prosody

Representation of speech in CorpAfroAs: Transcriptional strategies and prosodic units
Shlomo Izre'el and Amina Mettouchi

Tone and intonation
Bernard Caron

Part 2: Interfacing prosody, information structure and syntax

The intonation of topic and focus: Zaar (Nigeria), Tamasheq (Niger), Juba Arabic (South Sudan) and Tripoli Arabic (Libya)
Bernard Caron, Cécile Lux, Stefano Manfredi and Christophe Pereira

Quotative constructions and prosody in some Afroasiatic languages: Towards a typology
Il-Il Malibert and Martine Vanhove

Part 3: Cross-linguistic comparability

Glossing in Semitic languages: A comparison of Moroccan Arabic and Modern Hebrew
Angeles Vicente, Il-Il Malibert and Alexandrine Barantini

From the Leipzig Glossing Rules to the GE and RX lines
Bernard Comrie

Cross-linguistic comparability in CorpAfroAs
Amina Mettouchi, Graziano Savà and Mauro Tosco

Functional domains and cross-linguistic comparability
Zygmunt Frajzyngier and Amina Mettouchi

Part 4: Language contact

Language contact, borrowing and codeswitching
Stefano Manfredi, Marie-Claude Simeone-Senelle and Mauro Tosco

Part 5: Information technology

ELAN-Corpa: Lexicon-aided annotation in ELAN
Christian Chanard

Language index

Subject index

Berber

- Kabyle
- Tamasheq

Chadic

- Hausa
- Zaar

Creole (arabic)

- Juba

Cushitic

- Afar
- Beja
- Gawwada
- Tsamakko

OmotiC

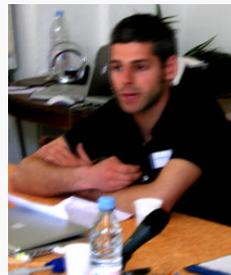
- Wolaytta

Semitic

- Moroccan Arabic
- Hebrew
- Tripolinian Arabic



The CorpAfroAs team



CorTypo (2013-2017)- <http://cortypo.huma-num.fr/>



HOME PROJECT LANGUAGES MEMBERS RESOURCES CONTACT US



CorTypo: Designing spoken corpora for cross-linguistic research

Financed by the Agence Nationale pour la Recherche (ANR), for 36 months (March 2013-March 2016). Prepared in 2011 and submitted in January 2012.

Principal Investigator: [Amina Mettouchi](#)

Directeur d'Etudes at [EPHE](#) (Ecole Pratique des Hautes Etudes), member of the CNRS laboratory [LLACAN](#)

Professional webpage (CV, publications): <http://aminamettouchi.linguanet.org>

Aim of the project

The aim of the CorTypo project is the elaboration of an innovative system of linguistic annotation of natural language corpora in lesser-described spoken

Innovative nature of the project

1. an annotation of sound-indexed texts that is based on the formal means existing in a given language, including prosodic means, linear orders, and

Latest News

Our next meeting will take place on
March 11, 12 and 13 2015.



The CorTypo team



Transcription & Annotation

ref@SP1 [419]	KAB_AM_CONV_01_SP1_212	KAB_AM_CONV_01	KAB_AM_CONV_01_SP1_214
tx@SP1 [418]	maɟfi ɔjəllis ɐ:: nəʕli //	237	a rʕəbbi qilillu //
mot@SP1 [1123]	mačči d jəllis ɐ:: n ʕli //	237	a rbbi qilillu
mb@SP1 [1260]	mačči d jəlli -s ɐ:: n aʕli //		a rbbi qilillu
ge@SP1 [1260]	NEG.ATTR COP daugh_KIN.3S HESIT GEN Aʕli //		VOC God flower
rx@SP1 [1260]	PTCL PRED N.KIN PRO HESIT PREP NP //		PTCL NP INTJ
ft@SP1 [238]	she was not Ali's daughter!		come on!
ref@SP2 [417]	KAB_AM_CONV_01_SP2_216	KAB_AM_CONV_01_SP2_217	KAB_AM_CONV_01_SP2_218
tx@SP2 [417]		ah ɔwəʔmas //	ɔwəʔmas //
mot@SP2 [1000]		ah d wəʔmas //	d wəʔmas
mb@SP2 [1061]		ah d wəʔma -s //	d wəʔma -s
ge@SP2 [1061]		INTJ COP sister KIN.3SG //	COP sister KIN.3SG
rx@SP2 [1061]		INTJ PRED N.KIN PRO //	PRED N.KIN PRO
ft@SP2 [239]		ah she was his sister!	his sister!
ref@SP3 [59]	KAB_AM_CONV_01_SP3_23		
tx@SP3 [59]			ih ɔjəllis nəʕli //
mot@SP3 [105]			ih d jəllis n
mb@SP3 [116]			ih d jəlli -s n
ge@SP3 [116]			yes COP dau_KIN_GEN
rx@SP3 [116]			ADV PRED N.KI PR PREP
ft@SP3 [27]			yes she's one of Ali's daughters

- Segmentation : Prosodic rather than syntactic because
 - Speech comes out in chunks
 - we need to segment continuous stretches of spontaneous data
 - Those chunks have a cognitive function
 - Chafe 1994: one new idea per IU, etc.
 - How is it possible to delimit clause boundaries anyway...?
 - Definitions of clauses are conceptual rather than formal
 - Tao 1996: : 'a verb plus its core arguments, with modifiers (e.g. locatives, adverbials, etc.) optionally present' (1996:17)
 - Ross 2011 : 'a grammatical construction which includes a predicate, its core arguments and adjuncts, where the predicate need not be verbal and may be adjectival or nominal' (2011:116)
 - ... nothing about where to place the left and right boundaries
 - especially in continuous stretches of spontaneous speech
 - ... even if ideal ('canonical') clauses can be delimited, what is the status of the 'residue' in the segmentation ?

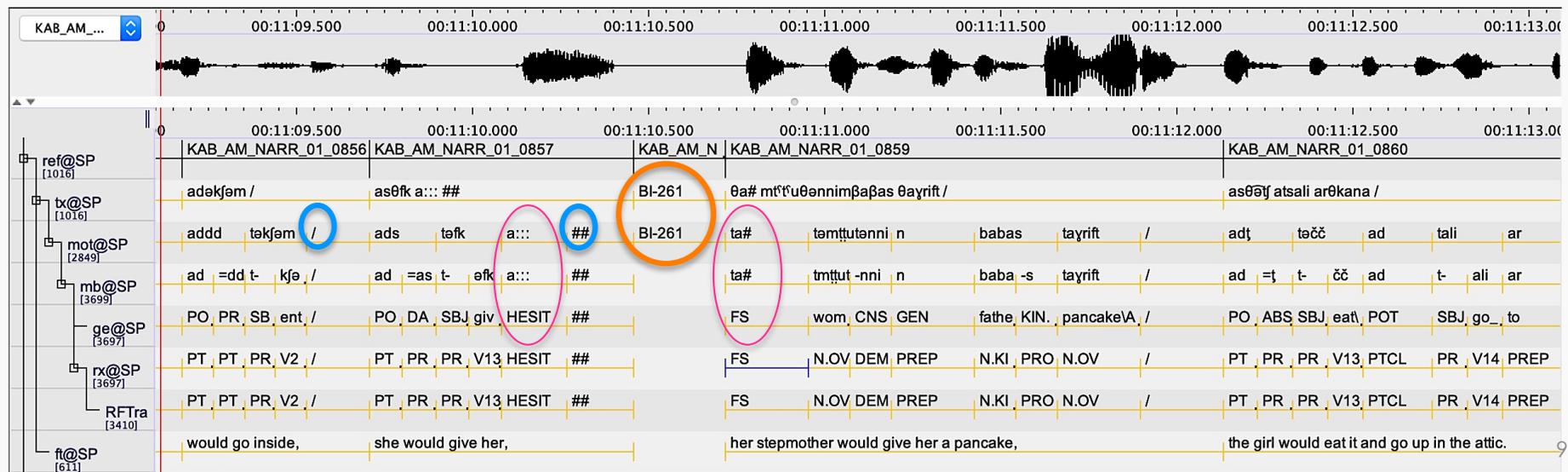
Systematic prosodic segmentation

- Which level of prosodic segmentation
 - Intonation Unit? ('Tone Unit', 'Tone group' ...)
 - single contour + boundaries (Cruttenden 1997 etc.)
 - Paratone?
 - high onset, brief pauses, gradual drift down in overall pitch height towards low ending (Brown 1992:92), continuing internal unit boundaries
 - Period?
 - speech stretch that shows declination along its paratones (« supradeclineation ») (Wichmann 2000:5.2.2), internal boundaries not necessarily continuing ones.
 - Interpausal? Random?

Izre'el, S. & A. Mettouchi. 2015. **Representation of speech in CorpAfroAs: Transcriptional strategies and prosodic units.** In: *Corpus-based Studies of lesser-described Languages: the CorpAfroAs Corpus of spoken AfroAsiatic*. Amina Mettouchi, Martine Vanhove & Dominique Caubet (eds.). Amsterdam, Philadelphia: John Benjamins (Studies in Corpus Linguistics 68), 13-41.

Granularity

- delimitation of pauses (boundaries at left and right of the pause) vs broad chunking (boundary in the middle of the pause)
- length (in ms.) and quality of pauses (breath intakes vs silent etc.)
- production phenomena (false starts, hesitations, etc.)



- « But those are only useful for pragmatic studies... »
 - absolutely not !
 - Prosodic segmentation and annotation of speech phenomena are crucial to define grammatical roles, and syntactic relations
- Aim of **CorpAfroAs** (create template amenable to automatic queries, involving prosodic segmentation et morphosyntactic annotation)
- Aim of **CorTypo** (constructions, their retrieval in corpora, cross-linguistic comparison of functions and functional domains)
- Underlying assumption: corpus annotation as mirroring/representing listener's input = at morpheme level, annotate everything that allows a form/function pairing, and only that (so that automatic queries can validate linguistic hypotheses) (avoid aprioristic annotation except for testing purposes)



THE DIRECT OBJECT IN KABYLE (BERBER)

Mettouchi, A. (2018) Prosodic segmentation and grammatical relations: the direct object in Kabyle (Berber).
REVISTA DE ESTUDOS DA LINGUAGEM, [S.l.], july 2018. ISSN 2237-2083.

Kabyle (Berber, Afroasiatic)



≈ 25 000 km²
≈ 5 million population



- Root & pattern morphology ; derivation
- Two genders and two numbers
 - marked on adjectives, on nouns, and on pronominal affixes and clitics hosted by verbs, nouns and prepositions.
- Two states (Absolute/Annexed), marked on nouns
- Minimal verbal predication = verb + bound personal pronoun(s)
 - In addition : noun phrases, prepositional phrases, adverbs.
- Clausal word-order = variable

The Direct Object in Kabyle

innajas zʿran wajθma θazʿarʿβiθ arjəmmanuza /

/ i-nna=as

zra-n

wajtma

tazərbit

SBJ.3SG.M-say\PFV=DAT.3SG

see\PFV-SBJ.3PL.M

brother\ANN.PL.M

carpet\ABS.SG.F

ar

jəmma

Nuza /

to

mother\ANN.SG.F

Nuza /

'he told him that his brothers had seen a carpet at Jemma Nuja's place'

(KAB_AM_NARR_02_505)

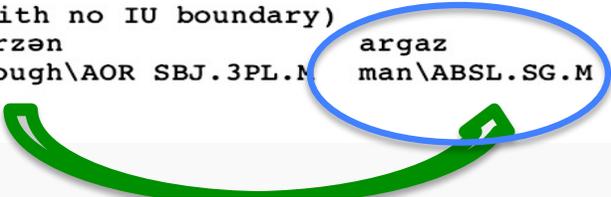


- Definition : a noun in the **absolute state (MORPH)**, **directly following the verb (SYNT) in the same intonation unit (PROSO)**, or possibly separated from it by a noun in the annexed state (=nominal subject), an adverb, a postverbal negator.

Importance of the prosodic boundary

- a noun in the absolute state (MORPH), directly following the verb (SYNT) ~~in the same intonation unit (PROSE)~~,

```
KAB_AM_NARR_03_0095 (transcription with no IU boundary)
ad  ruħən          ad  kərzən          argaz
POT go\AOR SBJ.3PL.M   POT plough\AOR SBJ.3PL.M   man\ABSL.SG.M
"they would go and plough a man"
```



```
KAB_AM_NARR_03_0096
ad  iṛuħ          ad  jawi          tajuga          ad  ikrəz          /
POT SBJ.3SG.M go\AOR   POT SBJ.3SG.M bring\AOR   pair_of_oxen\ABSL.SG.F   POT SBJ.3SG.M plough\AOR /
"he would go and bring a pair of oxen to plough"
```

```
KAB_AM_NARR_03_0095
ad  ruħən          ad  kərzən          / argaz
POT go\AOR SBJ.3PL.M   POT plough\AOR SBJ.3PL.M / man\ABSL.SG.M
"they would go and plough, one"
```

```
KAB_AM_NARR_03_0096
ad  iṛuħ          ad  jawi          tajuga          ad  ikrəz          /
POT SBJ.3SG.M go\AOR   POT SBJ.3SG.M bring\AOR   pair_of_oxen\ABSL.SG.F   POT SBJ.3SG.M plough\AOR /
would go and bring a pair of oxen to plough"
```



Counterexamples ?

- a noun in the absolute state (MORPH), directly following the verb (SYNT) ~~in the same intonation unit (PROSE)~~



Disfluencies:

- V [xxx] HESIT (/) Noun_{Abs}
- V [xxx] FS (## or /) Noun_{Abs}



awnəfkəy a::: / (KAB_AM_NARR_02_Midget_176)

adwən	fkəy	a:::	/
ad =wən	əfk	-ɣ	a:::
POT DAT.2PL.M	give\AORSBJ.1SG	HESIT	/
PTCLPRO	V13% PRO	HESIT	/

and I'll give you

ifrʌxa // (KAB_AM_NARR_02_Midget_177)

ifraxa		//
ifrax	-a	//
bird\ABSL.PL.MPROXa		//
N.OV	AFFX	//

those birds.

ixədmas θa:::# ## (KAB_AM_NARR_02_Midget_709)

ixədmas		ta:::# ##
i-	xdəm	=as ta:::# ##
SBJ.3SG.M	make\PFV	DAT.3SG FS ##
PRO	V23	PRO FS ##

he put on it

θabburθ duzzal / (KAB_AM_NARR_02_Midget_710)

tabburt	d	uzzal	/
tabburt	d	uzzal	/
door\ABSL.SG.F	COP	iron\ABSLM.SG	/
N.OV	PRED	N.OV	/

an iron door,

High rise :

- V [xxx] ↗ / Noun_{Abs}



alamma ʔəkkəsədd / ▶ (KAB_AM_NARR_01_0087)

alamma təkksədd /
 alamma t- əkks =dd /
 until SBJ.3SG.F take_away\PFV PROX /
 CONJ PRO V23 PTCL /
 until Clever Fatima manages to grab

faʕima ʔuħrʕiʃə / ▶ (KAB_AM_NARR_01_0088)

faʕima tuħrʕt /
 faʕima tuħrʕt /
 Faʕima clever /
 NP ADJ /

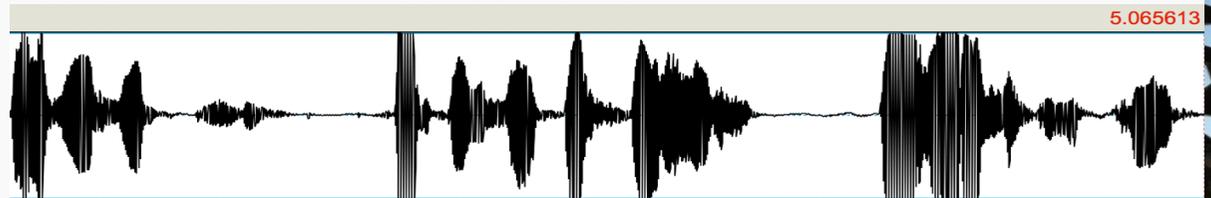
326 ▶ (KAB_AM_NARR_01_0089)

326

ayrum əgðəkkwan // ▶ (KAB_AM_NARR_01_0090)

ayrum g dəkkʰan //
 ayrum i dəkkʰan //
 bread\ABSL.SG.M LOC shelf\ANN.SG.M //
 N.OV PREP N.OV //

the bread on the shelf".



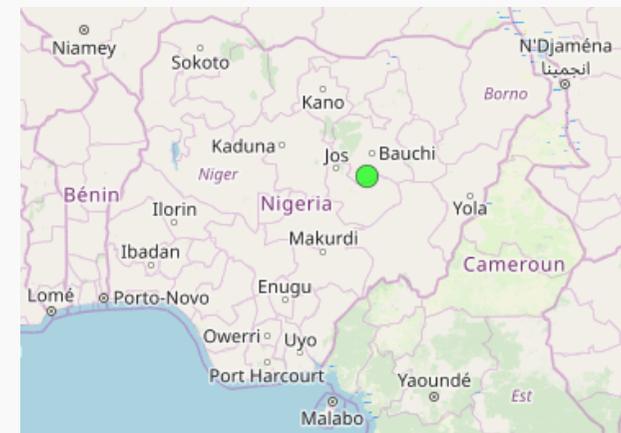
KAB_AM_NARR_01_0088	KAB_AM_NARR_01_0089	KAB_AM_N	KAB_AM_NARR_01_0091
alamma ʔəkkəsədd /	faʕima ʔuħrʕiʃə /	326	ayrum əgðəkkwan //
until she grabs	Clever Fatima	326	the bread on the shelf".
alamma təkksədd/	faʕima tuħrʕt/	326	ayrum g dəkkʰan//
5.065613			
Visible part 5.134476 seconds			5.13447
Total duration 5.134476 seconds			

‘Counterexamples’ = Supporting evidence

- Tight relationship of Direct Object with the verb is marked by the fact that if the noun in the **absolute state (MORPH)**, **directly following the verb (SYNT)** exceptionnally occurs **in the following intonation unit (PROSO)**
 - it is due to a **problem in speech production**, but the speaker signals to the hearer (through HESIT or FS) that the IU boundary is not to be interpreted as a regular boundary
 - it is an instance of **stylistic highlighting**, and the speaker signals to the hearer (through extra-high boundary tone) that the IU boundary is not to be interpreted as a regular boundary, and that the listener should consider the overall paratone CONTOUR for grammatical parsing

Centrality of Prosodic segmentation

- The **intonation boundary** is crucial to the definition
- Transcription and annotation of **dysfluencies** were crucial as supporting evidence for the formal definition of direct object in Kabyle
 - Annotation of main prosodic events (e.g. peaks) is also desirable (more difficult to achieve)
- **Prosodic chunking** (IU boundary) is a **formal means** contributing to the formal definition of Direct Object in Kabyle, at the same level as position relative to the verb (syntax), and absolute state (morphology)



REPORTED SPEECH IN FOUR AFROASIATIC LANGUAGES

Malibert, Il-H & Vanhove, Martine. 2015. **Quotative constructions and prosody in some Afroasiatic languages: Towards a typology.**

In: *Corpus-based Studies of lesser-described Languages: the CorpAfroAs Corpus of spoken AfroAsiatic*. Amina Mettouchi, Martine Vanhove & Dominique Caubet (eds.). Amsterdam, Philadelphia: John Benjamins (Studies in Corpus Linguistics 68), 117-169.



Reported speech in Afroasiatic

4 Afroasiatic languages

- ◆ Beja, North-Cushitic, SOV, only direct speech reports, no complementizer
- ◆ Zaar, Chadic, SVO, both direct (66%), and indirect (44%),
∃ complementizers *tu* or *wéj*
- ◆ Modern Hebrew, Semitic, SVO, both direct (30%), and indirect (70%),
∃ complementizers *se=* or *keilu* < *'like'
- ◆ Juba Arabic, Arabic based pidgin, SVO, both direct (95%), and indirect (5%),
marginal use of a complementizer *gale* < *'say'

Study

- Prosodic integration cline of speech reports within the quotative frame (cf. Genetti (2011) for Dolakha Newar)
 - “A number of features are used to mark discourse as direct speech, including the **relative positioning of prosodic and syntactic boundaries**, patterns of **terminal contours**, and changes in **loudness, pitch range, register, and timing**. As many of these features are scalar, direct speech reports can be placed on a **cline from prosodically independent to prosodically integrated** with respect to elements of the quotative frame. This variable prosodic behavior can be attributed to competition among discourse functional, syntactic, and production factors.” (Genetti 2011: 55)

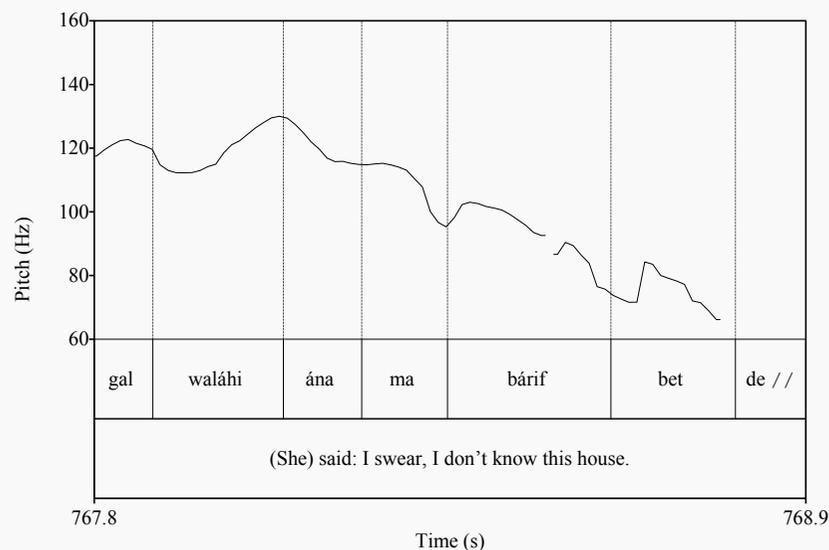
- Study limited to quotative frames with basic speech verbs, i.e. ‘say’ verbs, *di* (Beja), *wul* (Zaar), *wonosu*, *kelim*, and *gale* (Juba Arabic) *lomav* (Hebrew)
- Aim: first step for a typology of the interaction between prosody and reported speech
- If prosodic segmentation had not been included, none of the following generalizations would have been possible

- 1. If languages have *no complementizer*, the prosodic integration of speech reports within the same intonation unit as the quotative frame tends to be very high.

(Prosodic integration concerns the end of speech reports in SOV languages, and their onsets in SVO language).

- Juba Arabic 100% (marginal complementizer *gale*)
- Beja 90% (no complementizer) (10% due to dysfluencies)

PGA_SM_CONV_2



gále	waláhi	ána	ma	bi=árifu
say	by_god	1SG	NEG	IRR=know

bet	de	//
house	PROX.SG	



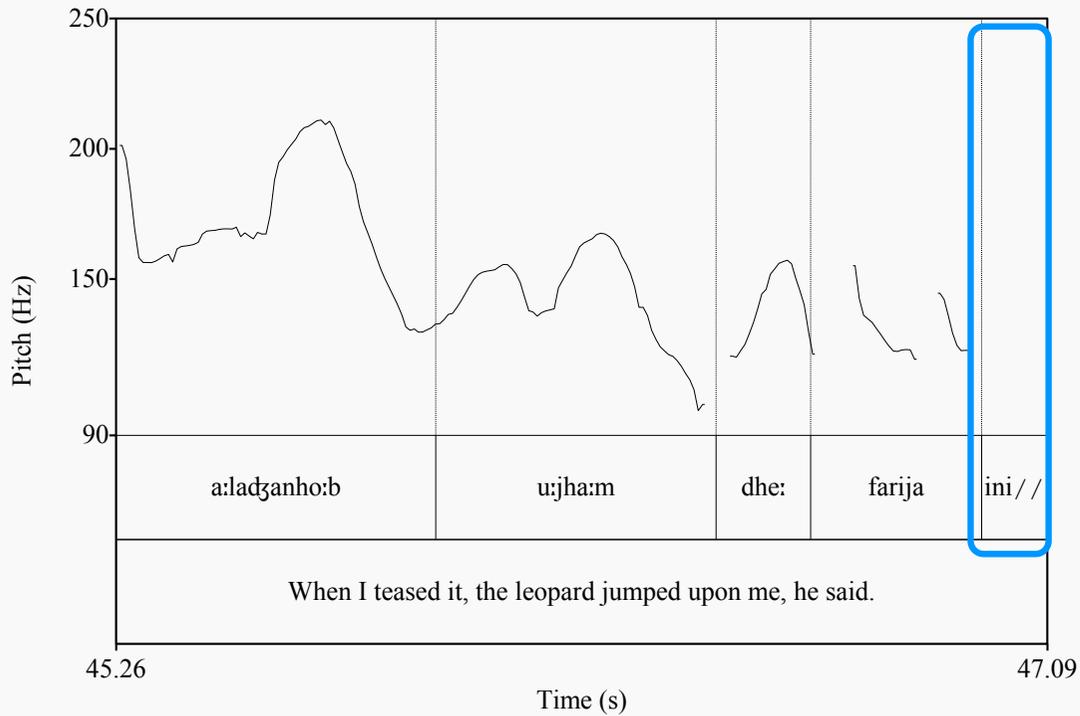
'(She) said: I swear, I don't know this house.'
 (PGA_SM_CONV_2_SP1_533) (Juba Arabic SVO)

Beja SOV

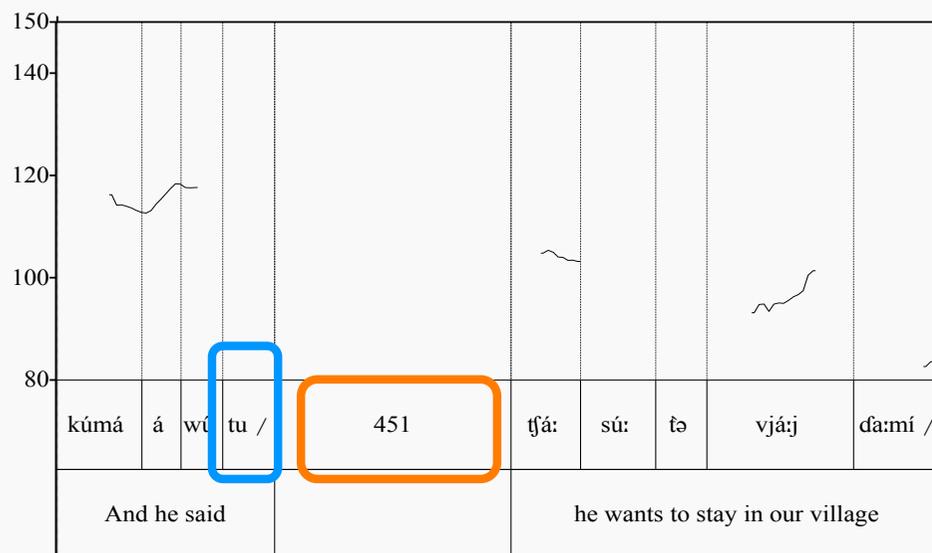
(3) *a:ladʒ-an = ho:b* *u: = jha:m* *d = he:*
tease-PFV.1SG=when DEF.SG.M.NOM=leopard DIR=1SG.ACC
far-ija *ini //*
jump-PFV.3SG.M say\PFV.3SG.M
‘When I teased it, the leopard jumped on me, he said.’
(BEJ_MV_NARR_15_leopard_051)



BEJ_MV_NARR_15_LEOPARD



- 2. Conversely, if languages *have a complementizer*, whatever their word orders, speech reports tend to be *less integrated* within the quotative frame.
 - Zaar 40% of speech reports NOT integrated within quotative frame
 - may be also Dolakha Newar, but Genetti (2011) does not provide any statistics.



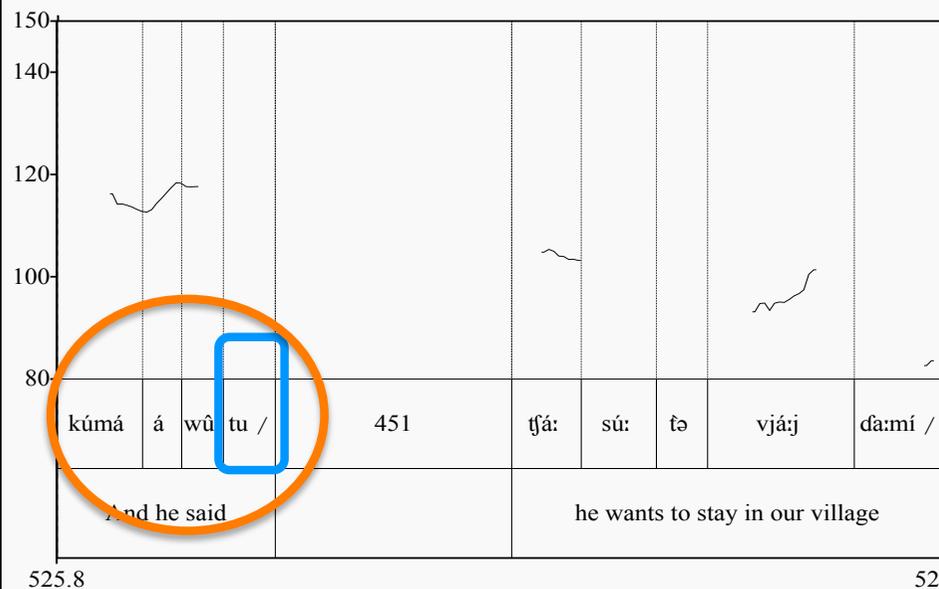
kúmá *á* *wul tu* / 451
 too 3SG.AOR say COMP
tʃá: *sú:* *tə* *vjá:-ú* *dá = mí* /
 3SG.IPFV want 3SG.SBV spend_day-DEF at=1PL.ACC
 ‘And he said that he wants to stay in our village’ (BC_SAY_Conv03
 (Boys)_SP1_556-558)

Zaar SVO



- 3. *Non-clitic complementizers* tend to be prosodically *integrated within the quotative frame* (and not within the reported discourse)

- Zaar: a prosodic boundary between the quotative frame and the speech report can only occur after the complementizer
- Juba Arabic



kú má *á* *wú* **tu / 451**
 too 3SG.AOR say COMP
tʃá: *sú:* *tə* *vjá:-ú:* *dɑ = mí* /
 3SG.IPFV want 3SG.SBV spend_day-DEF at=1PL.ACC
 ‘And he said that he wants to stay in our village’ (BC_SAY_Conv03
 (Boys)_SP1_556-558)

Zaar SVO



Generalizations leading to queries for BEJA

- A new IU is interpreted as reported speech iff it ends with the verb *di* 'say'

Look for / or // followed by X words followed by *SAY* (glossed V in rx) followed by / or //
– with **NO prosodic boundary in-between**

- is reported speech (sometimes only the end of the RS if spread over several IUs)
- Exceptions: out of 317 RS
 - in 40 cases the verb say is **detached in another IU**
 - onomatopoeia, exclamations, imperatives
 - in 12 cases the verb say is **in the middle of the RS**
 - between the main and the dependent clause
- when RS is long = several IUs (up until 13 consecutive IUs before the « say » verb)
 - look at pitch contours ? Other cues (higher pitch range and stronger intensity for whole passage)? Ideally, should be implemented in annotation

- in 40 cases the verb say is **detached in another IU**
 - onomatopoeia, exclamations, imperatives



alla:j bare:so:knaka / nhaska nija:ju / BI_541 indi e:n /
"God, more than yours his intention is cleaner", he says, they said;
"God has better intentions than yours!",he says, they said (BEJ_MV_NARR_08_drunkard_077-081)



- 4b. *In SVO languages it is the end of the speech report which is set off from the next IU.*
- 4a+4b may be good candidates for a universal prosodic cue of speech reports:
 - prosodic boundary at beginning of speech report for SOV
 - prosodic boundary at end of speech report for SVO
 - i.e. **a speech report is never in the same IU than the narrative part** (except for the quotative verb))
 - Boundary cues : strong preference for major terminal boundaries and pauses.
 - The rare cases with a minor boundary occur when the adjacent utterance is a dependent clause, as in Beja

Conclusion for reported speech

- Without prosodic segmentation, the typological hypotheses would not have been found
- Distinction between terminal and non-terminal boundaries allows for more fine-grained generalizations
- Prosody is essential to understand the relationship between narrative parts and reported speech in naturalistic data

SUBJECT VS TOPIC IN BEJA



M. Vanhove 24 June 2019. **Information structuring in Beja (North-Cushitic)**, 47th Annual Meeting of the North Atlantic Conference on Afroasiatic Linguistics (NACAL 47), INALCO, 24-26 June, 2019. Keynote lecture.

Subject vs topic

- Beja, North-Cushitic, SOV, topic markers extremely rare, three cases (NOM, ACC, GEN)
- Research question: does prosody help differentiate Subject from topics when no word change is involved, no topic marker is used and case marking is retained?
- Answer: yes it does!

Subject vs topic

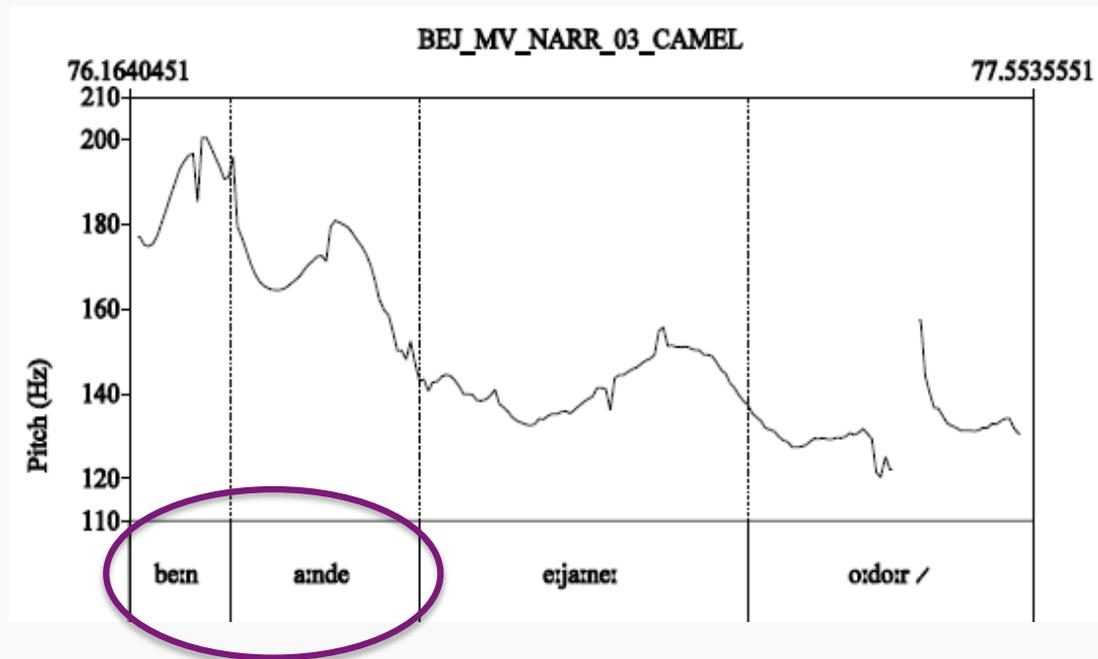
- Word order and intonation units (CorpAfroAs + CorTypo)
- N.nom = noun, noun phrases and pronouns in NOMINATIVE case

	$N_{\text{NOM}} + V = \text{within same IU}$	$N_{\text{NOM}} + V = \text{NOT within IU}$	Total
Nb N.nom	360	210	570
%	60%	34%	

Subject vs topic

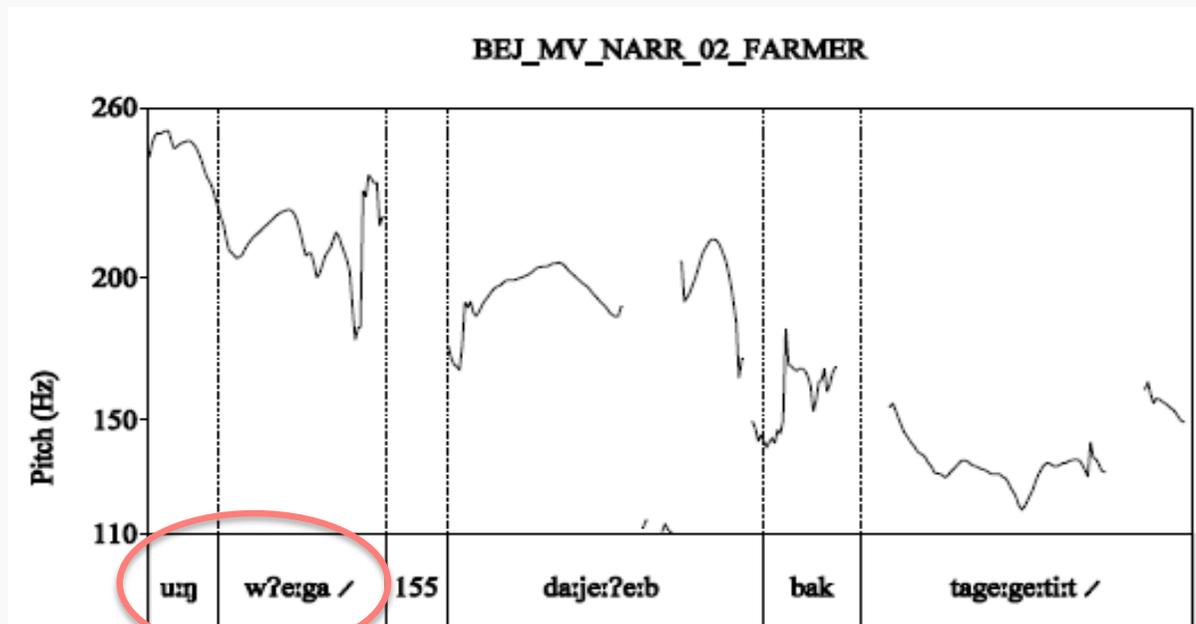
- No obvious pitch differences between topics and overt subjects
- No correlation between pitch and types of topics
 - BUT
- More **rising contours** with Topics
- More **falling contours** with Subjects
- The **cue** is in the **prosodic boundary**:
 - If **none** \Rightarrow **Subject**
 - If a **boundary** (with or without a pause) \Rightarrow **Topic**

In same IU = Subject



be:n a:=nda e:ja:-n=e:b o:=do:r
DIST.SG.M.NOM DEF.PL.M.NOM=man\PL come\PFV-3PL=REL.M DEF.SG.M.ACC=time
When these men came,

In different IU = Topic



u:n *wʔe:ga* / *155 da:je:b bak* *tage:g-e:ti:t* /
 PROX.SG.M.NOM DEF.SG.M=smoke . good thus be_high-CVB.SEQ .
 This smoke, after it is going well high like that...

(Smoke previously mentioned)

Conclusion

- Without prosodic segmentation it would have been difficult to discover the cue that distinguishes subjects from topics
- The same is true for objects vs topics

Synthesis and Conclusions

- Prosodic segmentation is relevant for core grammar
 - distinction minor-major essential
 - our results would not be reached with only // marked
- Fine-grained production phenomena are relevant for core grammar (dysfluencies etc.)
- Corpora as projections/representations of spoken performances should integrate the components of speech
 - probably also more precise prosodic features (F0 & I peaks, etc.)
 - probably also gestures
- This is not easy to implement (time-consuming)
 - evolving corpora (incremental transcription), based on quality data (recordings, videos, metadata, translations)



Thank you

*Don't sit and wait. Get out there, feel life.
Touch the sun, and immerse in the sea.
Jalal ad-Dīn Rūmī (1207-1273)*