

Corpus-Based Typology

(a qualitative/pilot approach)

Perspectives for cross-linguistic
comparison

Amina Mettouchi
EPHE, Paris & CNRS LLACAN
mettouchi@vjf.cnrs.fr

Talk dedicated to Bernard Comrie
MPI closing conference 1-3 May 2015



École Pratique
des Hautes Études

Langage, Langues et
LLACAN
Cultures d'Afrique Noire

Issues broached

- Corpora as language-specific resources
 - how to make them relevant for Typology
 - pilot projects (methodological)
- From CorpAfroAs to CorTypo
 - evolution of project-design
 - from phylum-internal cross-linguistic comparison
 - to empirical typology based on unrelated languages
- Corpus Annotation
 - A crucial issue for queries
 - Different annotations for different purposes
- Categories
 - Language-internal and cross-linguistic categories
 - comparative concepts, semantic maps...
 - the issue of similarity/comparability

Spontaneous Spoken Corpora

- Language-specific resources by nature
- Traditionally used for analysis of languages (+ elicitation, input for Grammars)
- Also within Documentation Projects
 - DOBES, ELDP
 - Attempts to use them for Cross-linguistic comparison / Typology
 - Verb/Noun project (PI- F. Seifart)
 - Referentiality project (PI- C. Lehmann)
- Dedicated pilot projects (for cross-linguistic comparison)
 - CorpAfroAs, CorTypo (PI- A. Mettouchi)
 - ...

- Annotation
 - in view of broad uses : CorpAfroAs, CorTypo
 - for a particular project: Referentiality, Verb-Nouns
- Differences
 - in granularity
 - in types of segmentation units
 - in number and types of tiers
 - in theoretical/methodological approach
 - ...


utterance_id [83]	FR0013									
utterance [83]	Coonj Heenaga hirohapa ru'ašnaga, wakaņšgejka giwā rehii!									
ft [83]	Hena first opens the window and (then) hollers for Froggy!									
utterance_to [402]	Coonj		Heenaga		hirohapa		ru'ašnaga,			
gramm_unit [593]	#	coonj	Heenaga	hirohapa	ru'ašnaga					
rp_gloss [593]	#	first	second.son-PROP	window=NR	open((SBJ.3SG)&(OBJ.3SG))					
unit_id [843]	gu1	gu2	gu3	gu4	gu5	pf1	pf2			
distr_cl [843]	#	adv	npr	nc	vt	pro_no	pro_no			
ref_ind [843]	-	-	002	010	-	002	010			
ref_typ [843]	-	-	spec	spec	-	spec	spec			
dd_se [843]	-	-	-	-	-	-	-			
sem_ro [843]	-	-	A	P	-	A	P			

Courtesy of
Iren Hartmann
Hoocak
(Reference
Project)

	5:03.800	00:05:04.000	00:05:04.200	00:05:04.400	00:05:04.600	00:05:04.800	00:05:05.000	00:05:05.200
ref@SP [1016]	KAB_AM_NARR_01_0376							
tx@SP [1016]	θənnajas anβədd anwali //							
mot@SP	tənnajas	ad	nβədd	ad	nwali	//		
mb@SP	t-	nna	=as	ad	n-	bədd	ad	n-
ge@SP	SBJ3	say\	DAT	POT	SBJ1PL	stand\A	POT	SBJ1PL
rx@SP	PRO	V13	PRO	PTCL	PRO	V24	PTCL	PRO
ft@SP [611]	she said "so that we can stand and see							

- <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

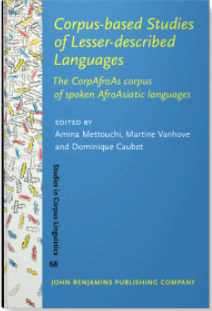
Corpus-based Studies of Lesser-described Languages


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](https://doi.org/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 Barontini

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 Frajzysgier 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

Participants



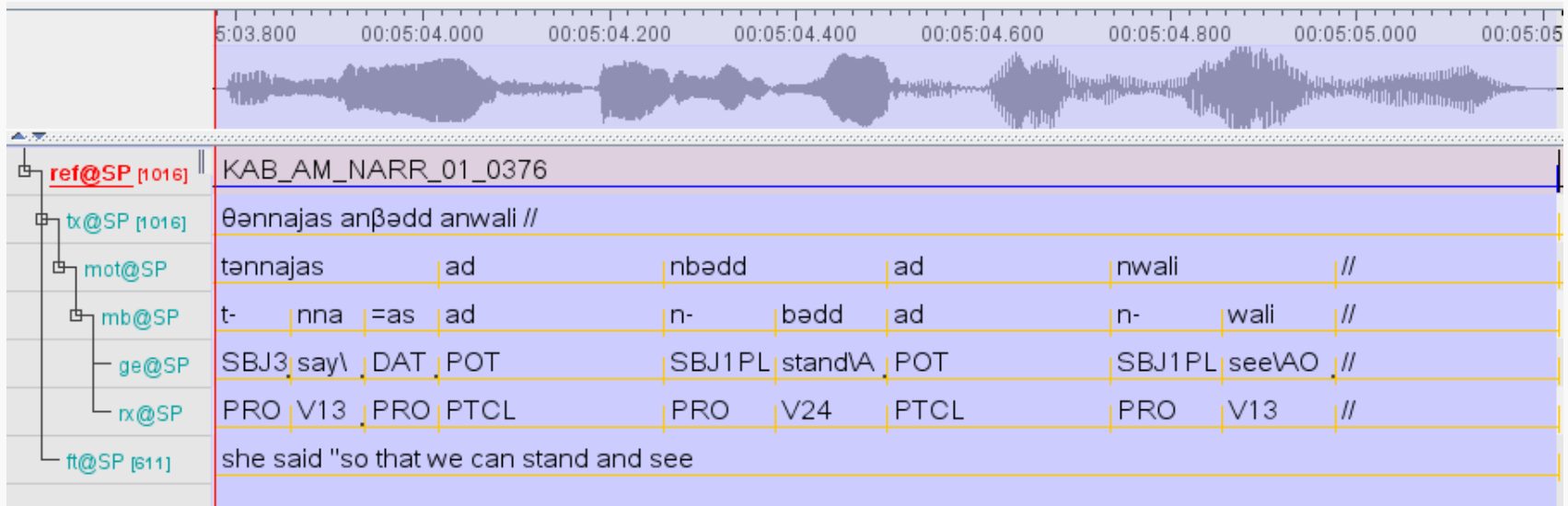
- **PI:** Amina Mettouchi
- **Engineering team:** Christian Chanard, Coralie Villes, Huyen-Tô Dan-Rabier
- **Experts:** Bernard Comrie, Shlomo Izre'el
- **Contributors:** Azeb Amha, Alexandrine Barontini, Bernard Caron, Cécile Lux, Il-Il Malibert-Yatsiv, Stefano Manfredi, Amina Mettouchi, Christophe Pereira, Mauro Tosco, Graziano Savà, Marie-Claude Simeone-Senelle, Martine Vanhove, Angeles Vicente.

CorpAfroAs

- Phylum-internal cross-linguistic comparison
 - 1 hour (40mn monological, 20mn dialogal) per language
 - 13 Afro-Asiatic languages
 - annotated on the same principles
 - same template for all the languages
 - functional annotation of morphemes
 - 1 morpheme = 1 gloss (ignore variation of readings)
 - » cf. IPFV regardless of progressive/habitual interpretation
 - 1 lexeme = 1 gloss
 - » cf. xdəm in KAB always 'make', regardless of contextual interpretation (make sthg vs. work)
 - part-of-speech information
 - free translation into English
 - same abbreviation for given gloss



Berber
<input type="checkbox"/> Kabyle
<input type="checkbox"/> Tamasheq
Chadic
<input type="checkbox"/> Hausa
<input type="checkbox"/> Zaar
Creole (arabic)
<input type="checkbox"/> Juba
Cushitic
<input type="checkbox"/> Afar
<input type="checkbox"/> Beja
<input type="checkbox"/> Gawwada
<input type="checkbox"/> Tsamakko
Omotic
<input type="checkbox"/> Wolaytta
Semitic
<input type="checkbox"/> Moroccan Arabic
<input type="checkbox"/> Hebrew
<input type="checkbox"/> Tripolinian Arabic



- ref** identifier for the annotation unit (time-associated)
- tx** transcription in broad phonetics into phonological words (SA)
- mot** intermediary tier with segmentation into morphosyntactic words (SS)
 - mb** morphophonological transcription into morphemes (SS)
 - ge** morpheme-by-morpheme gloss of mb according to the Leipzig Glossing Rules, expanded within the project (SA)
 - rx** part-of-speech and other information relevant for retrieval purposes (SA)
- ft** free translation into English (SA)

SA: symbolic association. SS: symbolic subdivision

TX : transcription

- segmentation into prosodic units
 - terminal vs. non-terminal boundaries
- transcription in ‘broad phonetics’, into phonological words
- conversations into separate actors

	KAB AM CONV 01 SP1 212	KAB AM CONV 01	KAB AM CONV 01 SP1 214
ref@SP1 [119]		237	
tx@SP1 [118]	maʃʃi ɔjallis o:: naʃli //		a rʻabbi qilillu //
mot@SP1 [1123]	maɕɕi d jallis o:: n ʃli //	237	a rbbi qilillu
mb@SP1 [1260]	maɕɕi d jalli -s o:: n aʃli //		a rbbi qilillu
ge@SP1 [1260]	NEG.ATTR COP daugh, KIN.3S HESIT GEN Aʃli //		VOC God flower
rx@SP1 [1260]	PTCL PRED N.KIN PRO HESIT PREP NP //		PTCL NP INTJ
ft@SP1 [1238]	she was not Ali's daughter!		come on!
ref@SP2 [117]			
tx@SP2 [117]		ah ɔwəɔmas //	ɔwəɔmas //
mot@SP2 [1000]		ah d wəɔmas //	d wəɔmas
mb@SP2 [1061]		ah d wəɔma -s //	d wəɔma -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 [1239]		ah she was his sister!	his sister !
ref@SP3 [59]			
tx@SP3 [59]			ih ɔjallis naʃli //
mot@SP3 [105]			ih d jallis n
mb@SP3 [116]			ih d jalli -s n
ge@SP3 [116]			yes COP dau KIN GEN
rx@SP3 [116]			ADV PRED N.KI PR PREP
ft@SP3 [127]			yes she's one of Ali's daughters

About TX

- No segmentation into syntactic units
- Phonological word : language-specific
- Possible queries given type of layout:
 - average length of Intonation Unit (ms)
 - number of phonological words per I.U.
 - length of pauses, speech/pause ratio...
 - frequency of some phonemes
 - ...
- comparison with MOT tier
(morphosyntactic words)

– TX

p^wintəd səβʃaθzəðmin gəsyarən /

– MOT

wwi-nt=dd sbʃa tzdmin n jsyarn /



(1)	<u>p^wintəd</u>	[səβʃ ^w aθzəðmin	<u>gəsy'arən</u>] /
	<u>wwi-nt=dd</u>	<u>sbʃa</u> <u>tzdmin</u>	n <u>jsyarn</u> /
	bring\PFV-SBJ3PL.F = PROX	seven bundle\ANN.PL.F	GEN firewood\ANN.PL.M /
	V14-PRO = PTCL	NUM N.OV	PREP N.OV /
	"they brought seven bundles of firewood"		

GE and RX in ELAN-CorpA

Dev. by C. Chanard (LLACAN)

- Annotation Lexicon (excerpt) →

- Glosses (excerpt)



CTP	Centripetal
CVB	<u>Converb</u>
DAT	Dative
DBL	Double
DECL	Declarative
DED	Deductive
DEF	Definite
DEICT	Deictic
DELAT	<u>Relative case</u>
DEO	Deontic modality
DEPMID	Deponent middle voice
DEPREC	Depreciative
DIFF	Diffusive
DIM	Diminutive
DIR	Directional
DIST	Distal
DISTR	Distributive
DITR	Ditransitive
DS	Different Subject
DTR	<u>Detransitivizer</u>
DU	Dual
DUB	Dubitative mood
DUR	Durative
ELAT	Elative case
EMPH	Emphatic

Lexicon Parse Parameters LinkedFile

... Nr	Lexicon	Variant	Gloss ▲	Tier X	Under...
1...	tarast		edible_herb\ABS.SG.F	N.cov	
1...	ṛəbbi		educate\AOR	V14	
1...	ṛəbbi		educate\IPFV	V14.PRFX.APHO	
1...	timəllalin		egg\ABS.F.PL	N.OV	
1...	həh		eh	INTJ	
1...	ja		either	CONJ	
7...	tmənzut		elder\ANN.SG.F	N.OV	
9...	trisiti		electricity\ABS	N.COVS	
9...	trisiti		electricity\ANN	N.COVS	
6...	taggara		end\ABS.F.SG	N.COVS	
2...	ann\tilat		enormous	PHRASE	
4...	kjəm	kəjm;...	enter\AOR	V23	
1...	kjam		enter\IPFV	V23	
4...	kjim		enter\NEGPFV	V23	
4...	kjəm		enter\PFV	V23	
8...	ssədhaj		entertain\CAUS.IPFV	V24.APHO	
65	irkəl		entirely	ADV	
5...	mənʃ		escape\PFV	V23	
1...	irumijən		European\ANN.PL.M	N.cov	
1...	<loro>		euros	N.CSW.FRA	
1...	<mème>		even	CONJ.CSW.FRA	
98	ula		even	ADV	
7...	yas		even_if	CONJ	
1...	taməddit		evening	ADV	
1...	tməddit		evening\ANN.SG.F	N.OV	
4...	kullej		everything	INDF.PRO	
1...	swaswa		exactly	ADV	
1...	tjili		exist\IPFV	V13%.PRFX.A...	
1...	lli		exist\NEGPFV	V13%	
1...	lla	lli	exist\PFV	V13%	
1...	ffəy	ffɣ	exit\AOR	V23	
6...	ssuffəy	ssuffɣ	exit\CAUS.AOR	V24	
3...	tjffəy		exit\IPFV	V23.PFX	
58	ffəy	ffɣ	exit\PFV	V23.TAM	

Glosses

- Consistent abbreviations
 - language-internal definitions

example: Absolute and Annexed states in Kabyle
(A. Mettouchi, working document)

Absolute(ABSL): ajoutée par Amina Mettouchi, le 12/12/2013.

In Kabyle, nouns appear under two forms, the Absolute state, and the Annexed state (see ANN).

The ABSOLUTE state is the default form of the noun, it has no function of its own, but is used each time the Annexed state is not used.

The absolute state is used in a number of constructions where it participates in the function of the construction, e.g. within a phrase, the sequence 'Head Noun'+ 'Noun in ABSL' codes adjectival modification.

See Mettouchi & Frajzyngier (2013) for more details:

<http://aminamettouchi.linguanet.org/wp-content/uploads/2012/06/LT-State-distinction-MettouchiFrajzyngier.pdf>

Annexed state (ANN): ajoutée par Amina Mettouchi, le 12/12/2013.

In Kabyle: nouns appear under two forms, the Absolute state (see ABSL), and the Annexed state.

The ANNEXED state is the form taken by the noun when it "provides the value (in the logical sense) for the variable in the function grammaticalized in a preceding constituent. A grammaticalized function is a function that has been encoded in the grammatical system of the language, through inflectional or syntactic means".

Mettouchi & Frajzyngier (2013:2):

<http://aminamettouchi.linguanet.org/wp-content/uploads/2012/06/LT-State-distinction-MettouchiFrajzyngier.pdf>

Queries

- For the corpus author:
 - test hypotheses
 - discover new constructions etc.
 - unmediated access to the corpus
- For the end-user
 - language specialist :
 - similar queries as author
 - mediation through list of glosses
 - phylum specialist, or typologist: necessary mediation
 - cross-linguistic queries
 - mediation through grammars or grammatical sketches + list of glosses

CTP	Centripetal
CVB	<u>Converb</u>
DAT	Dative
DBL	Double
DECL	Declarative
DED	Deductive
DEF	Definite
DEICT	Deictic
DELAT	<u>Relative case</u>
DEO	Deontic modality
DEPMID	Deponent middle voice
DEPREC	Depreciative
DIFF	Diffusive
DIM	Diminutive
DIR	Directional
DIST	Distal
DISTR	Distributive
DITR	Ditransitive
DS	Different Subject
DTR	<u>Detransitivizer</u>
DU	Dual
DUB	Dubitative mood
DUR	Durative
ELAT	Elative case
EMPH	Emphatic



Martine Vanhove

Beja grammatical sketch

Beja grammatical sketch

Martine Vanhove (LLACAN – CNRS, INALCO, PRES Sorbonne Paris-Cité)



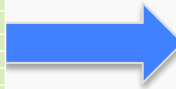
This sketch contains original material and analysis. The citation format is the following:
 Vanhove, Martine. 2014. Beja Grammatical Sketch. In Mettouchi, A. and C. Chanard
 (eds.) *The CorpaFroAs Corpus of Spoken AfroAsiatic Languages*. 68 pages. DOI:
<http://dx.doi.org/10.1075/sci.68.website>. Accessed on dd/mm/yyyy.
 Pages numbers should be referenced as they appear in this pdf.

Table of Contents

TABLE OF CONTENTS.....	1
1. INTRODUCTION	4
2. PHONOLOGY.....	5
2.1. CONSONANT PHONEMES	5
2.1.1. Allophones, free variation and phonological rules	6
2.2. VOCALIC PHONEMES.....	6
2.2.1. Allophones, free variation and phonological rules.....	7
2.3. PHONOTACTICS.....	7
2.4. STRESS.....	8
3. MORPHOLOGY.....	8
3.1. NOMINAL MORPHOLOGY.....	9
3.1.1. Noun categories.....	9
3.1.2. Noun patterns.....	9
3.1.2.1. "basic" noun patterns.....	9
3.1.2.2. Instrument Nouns.....	9
3.1.2.3. Place Nouns.....	10
3.1.3. Adjectives.....	10
3.1.3.1. Adjective patterns.....	10
3.1.3.2. Adjectivizer affix.....	10
3.1.4. Gender.....	11



OK	Language	Title	Author	Identifier	Size	Duration	Links
	Berber	Beja grammatical sketch	Martine VANHOVE	BEJ_MV_aGRAMMATICALSKETCH.PDF	1.3 Mb		
<input checked="" type="checkbox"/>	Kabyle	The shelter and the lovers	Martine VANHOVE	BEJ_MV_NARR_01_SHELTER	256 words	00:02:42	
<input checked="" type="checkbox"/>	Tamashq	The farmer and the djinn	Martine VANHOVE	BEJ_MV_NARR_02_FARMER	544 words	00:05:28	
<input checked="" type="checkbox"/>	Chadic	The camel race	Martine VANHOVE	BEJ_MV_NARR_03_CAMEL	392 words	00:03:49	
<input checked="" type="checkbox"/>	Hausa	The cat-djinn	Martine VANHOVE	BEJ_MV_NARR_04_DJINN	252 words	00:02:04	
<input checked="" type="checkbox"/>	Zaar	Lost in Eritrea	Martine VANHOVE	BEJ_MV_NARR_05_ERITREA	709 words	00:06:47	
<input checked="" type="checkbox"/>	Creole (arabic)	The foreigner, the Beja and the leopard	Martine VANHOVE	BEJ_MV_NARR_06_FOREIGNER	112 words	00:01:10	
<input checked="" type="checkbox"/>	Juba	Cold at the pilgrimage	Martine VANHOVE	BEJ_MV_NARR_07_COLD	136 words	00:01:16	
<input checked="" type="checkbox"/>	Cushitic	The drunkard who became a Muslim and a saint	Martine VANHOVE	BEJ_MV_NARR_08_DRUNKARD	358 words	00:03:25	
<input checked="" type="checkbox"/>	Afar	The jewel and the monster	Martine VANHOVE	BEJ_MV_NARR_09_JEWEL	111 words	00:01:04	
<input checked="" type="checkbox"/>	Beja	The rabbit and the camel	Martine VANHOVE	BEJ_MV_NARR_10_RABBIT	117 words	00:01:05	
<input checked="" type="checkbox"/>	Gawwada	Muna's coffee	Martine VANHOVE	BEJ_MV_NARR_11_COFFEE	98 words	00:00:33	
<input checked="" type="checkbox"/>	Tsamakko	The boy-eater witches	Martine VANHOVE	BEJ_MV_NARR_12_WITCH	289 words	00:02:22	
<input checked="" type="checkbox"/>	Omotiic	The dream by the grave	Martine VANHOVE	BEJ_MV_NARR_13_GRAVE	190 words	00:02:06	
<input checked="" type="checkbox"/>	Wolaytta	Sijadok the Christian and the Qadi	Martine VANHOVE	BEJ_MV_NARR_14_SIJADOK	552 words	00:05:31	
<input checked="" type="checkbox"/>	Semitiic	An old man kills a leopard	Martine VANHOVE	BEJ_MV_NARR_15_LEOPARD	197 words	00:01:51	
<input checked="" type="checkbox"/>	Moroccan Arabic	The Prophet, the fox and the crow	Martine VANHOVE	BEJ_MV_NARR_16_PROPHET_FOX	545 words	00:04:51	
<input checked="" type="checkbox"/>	Hebrew	The shoemaker and the fairies	Martine VANHOVE	BEJ_MV_NARR_17_SHOEMAKER	489 words	00:04:27	
<input checked="" type="checkbox"/>	Tripolitanian Arabic	Adam and the devil	Martine VANHOVE	BEJ_MV_NARR_18_ADAM_DEVIL	583 words	00:05:26	
	Hausa	Hausa grammatical sketch	Bernard CARON	HAU_BC_aGRAMMATICALSKETCH.PDF	563 Kb		
<input checked="" type="checkbox"/>	Hausa	Hausa Conversation 1 (Boys)	Bernard CARON	HAU_BC_CONV_01	2410 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Conversation 2 (Boys)	Bernard CARON	HAU_BC_CONV_02	1638 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Conversation 3 (Girls)	Bernard CARON	HAU_BC_CONV_03	1326 words		
<input checked="" type="checkbox"/>	Hausa	Hausa conversation 4 (Men)	Bernard CARON	HAU_BC_CONV_04	4990 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Narration 1 (Women)	Bernard CARON	HAU_BC_NARR_01	860 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Narration 2 (Women)	Bernard CARON	HAU_BC_NARR_02	611 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Narration 3 (Girls)	Bernard CARON	HAU_BC_NARR_03	146 words		



Search Domain
97 file(s)

Concordances and Lists
List of tier type: mot | ? order: alphabetically by frequency (LIST)
Concordances in: mot | ? CONCORDANCE Context length (chars): 50

Search
case sensitive: regular expression:
Minimal duration (ms): 0 Maximal duration (ms):
Left Context Target Right Context (Clear)
NEG.EXS All Tiers: ?
NEG.EXS All Tiers: ?
NEG.EXS All Tiers: ?
NEG.EXS All Tiers: ?
FIND



- HEB_IM_NARR_4.EAF: NEG.EXS { HEB_IM_NARR_4.SP1_326 }
- HEB_IM_NARR_7.EAF: NEG.EXS { HEB_IM_NARR_7.SP1_0206 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0209 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0245 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0250 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0254 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0256 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0261 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0381 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0586 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0597 }
 NEG.EXS { HEB_IM_NARR_7.SP1_0612 }
 NEG.EXS { HEB_IM_NARR_7.SP2_050 }
 NEG.EXS { HEB_IM_NARR_7.SP2_083 }
 NEG.EXS { HEB_IM_NARR_7.SP2_087 }
 NEG.EXS { HEB_IM_NARR_7.SP2_163 }
 NEG.EXS { HEB_IM_NARR_7.SP2_207 }
 NEG.EXS { HEB_IM_NARR_7.SP2_315 }
 NEG.EXS { HEB_IM_NARR_7.SP2_325 }
- KAB_AM_NARR_01.EAF: NEG.EXS { KAB_AM_NARR_01_0304 }
 NEG.EXS { KAB_AM_NARR_01_0321 }
 NEG.EXS { KAB_AM_NARR_01_0574 }
 NEG.EXS { KAB_AM_NARR_01_0653 }
 NEG.EXS { KAB_AM_NARR_01_0779 }
 NEG.EXS { KAB_AM_NARR_01_0783 }
 NEG.EXS { KAB_AM_NARR_01_0901 }
- KAB_AM_NARR_02.EAF: NEG.EXS { KAB_AM_NARR_02_510 }
 NEG.EXS { KAB_AM_NARR_02_556 }
- KAB_AM_NARR_03.EAF:

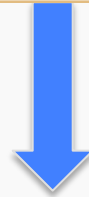
Extend Display

ulaf // (KAB_AM_NARR_01_0304)
ulaf //
ulaf //
NEG.EXS //
PRED //
there's nothing left."



say\PFV.SBJ.3PL.F.DAT.3SG so now HESIT / V13% PRO PRO CONJ ADV HESIT /
They told their little sister, "now,
459 (KAB_AM_NARR_01_0302)
459
afaranɔf // (KAB_AM_NARR_01_0303)
afu ara nəçč //
afu ara n- çč //
what REL.IRR SBJ.1PL.eat\AOR //
INTERR.PRO N.INDF PRO V13% Q
what are we going to eat ?
ulaf // (KAB_AM_NARR_01_0304)
ulaf //
ulaf //
NEG.EXS //
PRED //
there's nothing left."
BI-338 (KAB_AM_NARR_01_0305)
BI-338
nnantas anɔf danɔf fat'ima tuhrift // (KAB_AM_NARR_01_0306)
nnantas ad nəf'di ad nəçč fat'ima tuhrift //
nna -nt =as ad n- f'di ad n- çč fat'ima tuhrift //
say\PFV.SBJ.3PL.F.DAT.3SG POT SBJ.1PL.pass\AOR POT SBJ.1PL.eat\AOR Fat'ima clever //
V13% PRO PRO PTCL PRO V13% PTCL PRO V13% NP ADJ //
They said "let's eat Clever Fatima !"

OK	Language	Title	Author	Identifier	Size	Duration	Links
	Berber						
<input checked="" type="checkbox"/>	Kabyle						
<input checked="" type="checkbox"/>	Tamasheq						
	Chadic						
<input checked="" type="checkbox"/>	Hausa						
<input checked="" type="checkbox"/>	Zaar						
	Creole (arabic)						
<input checked="" type="checkbox"/>	Juba						
	Cushitic						
<input type="checkbox"/>	Afar						
<input checked="" type="checkbox"/>	Beja						
<input type="checkbox"/>	Gawwada						
<input checked="" type="checkbox"/>	Tsamakko						
	Omotic						
<input type="checkbox"/>	Wolaytta						
	Semitic						
<input type="checkbox"/>	Moroccan Arabic						
<input type="checkbox"/>	Hebrew						
<input type="checkbox"/>	Tripolinian Arabic						
41	Beja	Beja grammatical sketch	Martine VANHOVE	BEJ_MV_aGRAMMATICALSKETCH.PDF	1.3 Mb		
<input checked="" type="checkbox"/>	Beja	The shelter and the lovers	Martine VANHOVE	BEJ_MV_NARR_01_SHELTER	256 words	00:02:42	
<input checked="" type="checkbox"/>	Beja	The farmer and the djinn	Martine VANHOVE	BEJ_MV_NARR_02_FARMER	544 words	00:05:28	
<input checked="" type="checkbox"/>	Beja	The camel race	Martine VANHOVE	BEJ_MV_NARR_03_CAMEL	392 words	00:03:49	
<input checked="" type="checkbox"/>	Beja	The cat-djinn	Martine VANHOVE	BEJ_MV_NARR_04_DJINN	252 words	00:02:04	
<input checked="" type="checkbox"/>	Beja	Lost in Eritrea	Martine VANHOVE	BEJ_MV_NARR_05_ERITREA	709 words	00:06:47	
<input checked="" type="checkbox"/>	Beja	The foreigner, the Beja and the leopard	Martine VANHOVE	BEJ_MV_NARR_06_FOREIGNER	112 words	00:01:10	
<input checked="" type="checkbox"/>	Beja	Cold at the pilgrimage	Martine VANHOVE	BEJ_MV_NARR_07_COLD	136 words	00:01:16	
<input checked="" type="checkbox"/>	Beja	The drunkard who became a Muslim and a saint	Martine VANHOVE	BEJ_MV_NARR_08_DRUNKARD	358 words	00:03:25	
<input checked="" type="checkbox"/>	Beja	The jewel and the monster	Martine VANHOVE	BEJ_MV_NARR_09_JEWEL	111 words	00:01:04	
<input checked="" type="checkbox"/>	Beja	The rabbit and the camel	Martine VANHOVE	BEJ_MV_NARR_10_RABBIT	117 words	00:01:05	
<input checked="" type="checkbox"/>	Beja	Muna's coffee	Martine VANHOVE	BEJ_MV_NARR_11_COFFEE	58 words	00:00:33	
<input checked="" type="checkbox"/>	Beja	The boy-eater witches	Martine VANHOVE	BEJ_MV_NARR_12_WITCH	289 words	00:02:22	
<input checked="" type="checkbox"/>	Beja	The dream by the grave	Martine VANHOVE	BEJ_MV_NARR_13_GRAVE	190 words	00:02:06	
<input checked="" type="checkbox"/>	Beja	Sijadok the Christian and the Qadi	Martine VANHOVE	BEJ_MV_NARR_14_SIADOK	552 words	00:05:31	
<input checked="" type="checkbox"/>	Beja	An old man kills a leopard	Martine VANHOVE	BEJ_MV_NARR_15_LEOPARD	197 words	00:01:51	
<input checked="" type="checkbox"/>	Beja	The Prophet, the fox and the crow	Martine VANHOVE	BEJ_MV_NARR_16_PROPHET_FOX	545 words	00:04:51	
<input checked="" type="checkbox"/>	Beja	The shoemaker and the fairies	Martine VANHOVE	BEJ_MV_NARR_17_SHOEMAKER	489 words	00:04:27	
<input checked="" type="checkbox"/>	Beja	Adam and the devil	Martine VANHOVE	BEJ_MV_NARR_18_ADAM_DEVIL	583 words	00:05:26	
63	Hausa	Hausa grammatical sketch	Bernard CARON	HAU_BC_aGRAMMATICALSKETCH.PDF	563 Kb		
<input checked="" type="checkbox"/>	Hausa	Hausa Conversation 1 (Boys)	Bernard CARON	HAU_BC_CONV_01	2410 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Conversation 2 (Boys)	Bernard CARON	HAU_BC_CONV_02	1638 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Conversation 3 (Girls)	Bernard CARON	HAU_BC_CONV_03	1326 words		
<input checked="" type="checkbox"/>	Hausa	Hausa conversation 4 (Men)	Bernard CARON	HAU_BC_CONV_04	4990 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Narration 1 (Women)	Bernard CARON	HAU_BC_NARR_01	860 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Narration 2 (Women)	Bernard CARON	HAU_BC_NARR_02	611 words		
<input checked="" type="checkbox"/>	Hausa	Hausa Narration 3 (Girls)	Bernard CARON	HAU_BC_NARR_03	146 words		



Search Domain

97 file(s)

Concordances and Lists

List of tier type [?](#) order alphabetically by frequency

Concordances in [?](#) Context length (chars) :

Search

[?](#)

Minimal duration (ms): Maximal duration (ms):

Left Context	Target	Right Context	
<input type="text"/>	<input type="text" value="NEG.EXS"/>	<input type="text"/>	<input type="text" value="All Tiers"/> ?
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text" value="All Tiers"/> ?
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text" value="All Tiers"/> ?
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text" value="All Tiers"/> ?





Extend Display

ulaf // (KAB_AM_NARR_01_0304)
 ulaf //
 ulaf //
 NEG.EXS //
 PRED //
 there's nothing left."



say\PFV SBJ.3PL.FDAT.3SG so now HESIT /
 V13% PRO PRO CONJ ADV HESIT /
 They told their little sister, "now,

459 (KAB_AM_NARR_01_0302)
 459

afaranəfj // (KAB_AM_NARR_01_0303)
 afu ara nəčč //
 afu ara n- čč //
 what REL.IRR SBJ.1PL eat\AOR //
 INTERR.PRO N.INDF PRO V13% Q
 what are we going to eat ?

ulaf // (KAB_AM_NARR_01_0304)
 ulaf //
 ulaf //
 NEG.EXS //
 PRED //
 there's nothing left."

BI-338 (KAB_AM_NARR_01_0305)
 BI-338

nnantas anəʕdanəfj fatʕima θuħrijft // (KAB_AM_NARR_01_0306)
 nnantas ad nəʕdi ad nəčč faʕima tuħrijft //
 nna -nt =as ad n- ʕdi ad n- čč faʕima tuħrijft //
 say\PFV SBJ.3PL.FDAT.3SG POT SBJ.1PL pass\AOR POT SBJ.1PL eat\AOR Faʕima clever //
 V13% PRO PRO PTCL PRO V13% PTCL PRO V13% NP ADJ //
 They said "let's eat Clever Fatima !"

- **HEB_IM_NARR_4.EAF :**
 - NEG.EXS { [HEB_IM_NARR_4_SP1_326](#) }
- **HEB_IM_NARR_7.EAF :**
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0206](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0209](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0245](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0250](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0254](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0256](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0261](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0381](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0586](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0597](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP1_0612](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP2_050](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP2_083](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP2_087](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP2_163](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP2_207](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP2_315](#) }
 - NEG.EXS { [HEB_IM_NARR_7_SP2_325](#) }
- **KAB_AM_NARR_01.EAF :**
 - NEG.EXS { [KAB_AM_NARR_01_0304](#) }
 - NEG.EXS { [KAB_AM_NARR_01_0321](#) }
 - NEG.EXS { [KAB_AM_NARR_01_0574](#) }
 - NEG.EXS { [KAB_AM_NARR_01_0653](#) }
 - NEG.EXS { [KAB_AM_NARR_01_0779](#) }
 - NEG.EXS { [KAB_AM_NARR_01_0783](#) }
 - NEG.EXS { [KAB_AM_NARR_01_0901](#) }
- **KAB_AM_NARR_02.EAF :**
 - NEG.EXS { [KAB_AM_NARR_02_510](#) }
 - NEG.EXS { [KAB_AM_NARR_02_556](#) }
- **KAB_AM_NARR_03.EAF :**

Cross-Linguistic Comparison in CorpAfroAs

- Directional extensions (Mettouchi, Savà & Tosco 2015), attach to verbs
 - Hausa, Zaar (Chadic), Kabyle, Tamasheq (Berber), Gawwada, Ts'amakko (Cushitic)
 - Grammatical sketches & Lists of Glosses
 - DIR (directional, precisely 'ventive') in Hausa, CTP (centripetal) in Zaar
 - PROX (proximal) & DIST (distal) in Kabyle & Tamasheq
 - ASS (assertive) & DAT(dative) in Ts'amakko and CTP (centripetal) & CFG(centrifugal) in Gawwada

Examples

dà mukà ta:so: / ▶ (HAU_BC_CONV_01_SP2_306)

dà	mukà	ta:so:	/
dà	mukà	ta:so:	/
as	1PL.PFV.FOC	leave.DIR	/
▶ CONJ/hom	PNG.TAM	V6	/

As we had left,

munà: sa:ran ità:tfe: munà: sa:ran ità:tfe: / ▶ (HAU_BC_CONV_01_SP2_307)

munà:	sa:ran	ità:tfe:	munà:	sa:ran	ità:tfe: /
munà:	sa:ra: -n	ità:tfe:	munà:	sa:ra: -n	ità:tfe: /
1PL.CONT.NFOC	felling-POSL	wood	1PL.CONT.NFOC	felling-POSL	wood /
PNG.TAM	N.V1 -PTCL.SYNT	N	PNG.TAM	N.V1 -PTCL.SYNT	N /

we were cutting wood,

antr^huñ arɟɟixiw / ▶ (KAB_AM_NARR_03_0478)

adn	=n	t-	ruñ	ar	ɟɟixiw	/
ad	=n	t-	ruñ	ar	ɟɟix	-iw /
POT	DIST	SBJ3SG.F	go\AOR	to	teacher\ANN.SG.M	POSS1SG /
▶ PTCL	PTCL	PRO	V24	PREP	N.COV	PRO /

she would go to the teacher,

asθini ə::: / ▶ (KAB_AM_NARR_03_0479)

ads	=as	t-	tini	ə:::	/
ad	=as	t-	ini	ə:::	/
POT	DAT3SG	SBJ3SG.F	say\AOR	HESIT	/
PTCL	PRO	PRO	V13%	HESIT	/

she would tell him

Proportions

- Hausa and Zaar : only one directional extension (Hi tone +-o; -di)
- Kabyle and Tamasheq: two extensions
 - Tamasheq: DIST= 40%, PROX=60%
 - Kabyle: DIST= 0,1%, PROX= 99,9%
 - same diachronic origin =in/=n vs. =du /dd
- Ts'amakko and Gawwada:
 - Ts'amakko: DAT= 3%; ASS= 97%
 - Gawwada: CTP= 21%; CFG= 79%
 - same diachronic origin =nu vs. =na

Main types of verbs

- Hausa and Zaar
 - motion and handling verbs,
- Kabyle
 - verbs of motion, position and handling,
 - verbs of saying
 - all sorts of other verbs except statives
- Tamasheq
 - motion verbs and verbs of saying
- Gawwada and Ts'amakko
 - verbs of motion and handling,
 - verbs of saying
 - other verbs (human activities)

Corpus queries as suggestions, preliminary investigations for more detailed cross-linguistic/typological research.



- Grammaticalisation path from motion verbs
 - ? diffusion to potentially all types of dynamic verbs (cf. Kabyle) – with modal values, via:
 - handling verbs
 - verbs of saying
- Link between deictic motion and speaker-stance?
 - movement towards speaker/addressee in Tamasheq, and Gawwada
 - + viewpoint of the speaker/ addressee in Kabyle
- etc.

Achievements and Limitations

- Achievements
 - 13 hours of prosodically-segmented, and morphosyntactically-annotated spoken data in 13 lesser-described languages
 - common annotation scheme & template
 - 40% dialogal, 60% monological
- Limitations
 - No clear-cut distinction between language-internal categories & cross-linguistic ones
 - Necessary mediation through grammatical sketches, grammars...

CorTypo (2013-2017)

[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.



Participants



- **PI:** Amina Mettouchi
- **Engineering team:** Christian Chanard, Mourad Aouini, Tahar Meddour
- **Scientific Managers:** Zygmunt Frajzyngier (Functional database), Bernard Comrie (Table of Categories), Martine Vanhove (Corpus)
- **Contributors:** Evangelia Adamou, Azeb Amha, Isabelle Bril, Bernard Caron, Denis Creissels, Zygmunt Frajzyngier, Katharina Haude, Il-Il Malibert-Yatsiv, Stefano Manfredi, Amina Mettouchi, Nicolas Quint, Stéphane Robert, Paulette Roulon-Doko, Yvonne Treis, Martine Vanhove.
- **Administrative coordinator:** Jeanne Zerner
- **Invited IT expert:** Marc Kemps-Snijders

- <http://cortypo.humanum.fr/>
- 14 languages (different phyla)
- A pilot corpus, fully L-annotated
 - same template as CorpAfroAs,
 - with additional functionalities in Elan (allowing retrieval of complex constructions & reference-tracking)

Movima (Amerindian isolate)
Ixcatec (Oto-Manguean)
Amis (Austronesian)
Gbaya (Ubangian)
Wolof (Atlantic)
Koalib (Heiban)
Kabyle (Berber)
Wandala (Chadic)
Zaar (Chadic)
Beja (Cushitic)
Kambaata (Cushitic)
Wolaitta (Omotic)
Baggara Arabic (Semitic)
Hebrew (Semitic)

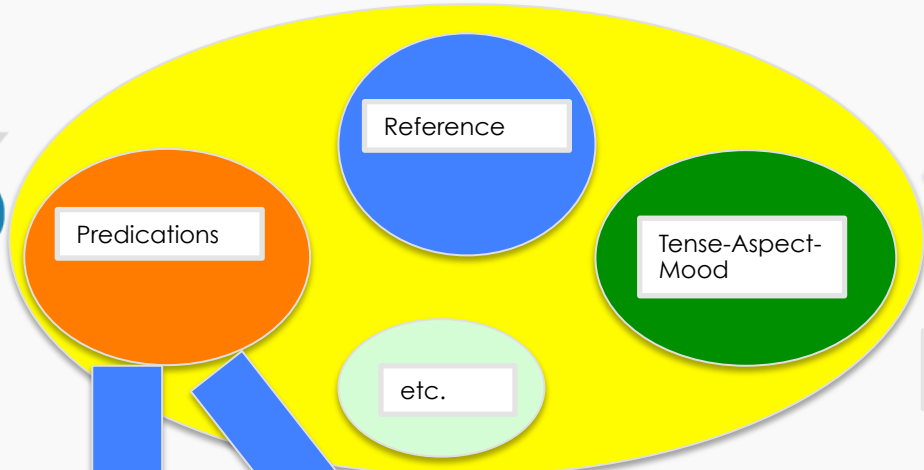


- Deliverables
 - A pilot corpus
 - Interfaced with a comparative database
 - And a table of categories (gloss, abbreviation and full language-internal definitions of categories + cross-linguistic definitions (supervised by B. Comrie))
- The comparative database
 - bottom-up empirically-built functional domains
 - inspired by Frajzyngier's approach (Frajzyngier & Mycielski 1998, Frajzyngier & Shay 2003, Frajzyngier 2013, Frajzyngier & Mettouchi 2015, ...)
 - comparison among Functional domains
 - « The functional domain is a set of functions that all share one semantic characteristic, and the forms that realize them are in complementary distribution within a relevant constituent.” (Z.F.)



Similarities and differences (F&M 2015)

- Languages are similar/different
 - Different functional domains encoded in the grammar
 - L1 Aspect, L2 Tense, L3 Aspect+Tense...
 - Different structures of a given F.D.
 - L1 Three Aspects, L2 Five Aspects ...
 - Different grammaticalized meanings
 - Perfective in a language with a Negative Perfective (Kabyle) is different from Perfective in a language with no Negative Perfective (Shilha, Siwi...)



current query interface

Langue	Domaine	Predication	Construction
Kabyle	PRED + NP	Affecting Subject	Affecting subject predication has the form prefix s- (s- before vowel)+ verb
		Appreciative	Appreciative predication has the form a]k + absolutive clitic or f]wa + absolutive clitic Look for be_beautiful / great in ge and PRED in rx, at a distance of -1 "ge" cell of ABSV in ge and PRO in rx.
		Ascriptive	Ascriptive predication has the forms: a) predicative copula d + adjective or noun b) predicative copula d + adjective iri "bad" or f]ali "good" + absolutive clitic OR predicative copula d + adjective iri "bad" or f]ali "good" +NP
			a) COP in ge and PRED in rx, at a distance of -1 "ge" cell of ADJ in rx. b) COP in ge and PRED in rx, at a distance of -1 "ge" cell of good (or bad) in ge, itself at a distance of -1 "ge" cell of ABSV in ge and PRO in rx
		Dynamic	Dynamic predication has the form : labile verb+obligatory subject affix, plus either an absolutive clitic pronoun, or a noun in the absolute state.
		Externally-affected subject	Externally-affected subject has the form tw-prefix + verb root + subject affix
		Internally affected subject	Internally affected subject has the form mm- or nn- prefix + verb root + subject affix
		Negative Ascriptive	Negative ascriptive predication has the form mat]fi + COP + adjective or noun
		Negative existential	Negative Existential has the form ulaf, preceded or followed by an NP in the absolute state, where the form ulaf is the negative existential predicate.
		Negative locative	Negative locative predication has the form ulaf+absolutive clitic, where the form ulaf is the negative existential predicate. The referent of the absolutive clitic can be expressed either by an NP in the annexed state following the predicate+clitic, or by an NP in the absolute state preceding the predicate+clitic ulaf + absolutive pronoun

Kambaata	PRED + NP	Ascriptive	The subject is nominative-marked, the property is either expressed by an adjective or a relativised verb to which the equational/ascriptive copula is suffixed.
		Causative	Morphological derivation, suffix -(i)s (the suffix sometimes merges with the stem-final consonant) or -(i)siis; the distribution of -(i)s vs. -(i)siis is partly grammatically, partly lexically determined and thus not entirely predictable; the causer/controller is marked by the NOM case, the causee is marked by the ACC or DAT case.
		Distributive	Morphological derivation consisting of three morphemes, the middle morpheme, the passive morpheme and the causative morpheme
		Equational	The subject is nominative-marked, the class/group to which the referent belongs is expressed by a noun to which the equational/ascriptive copula (COP2) is suffixed.
		Existential	Existential/copular verb yoo- (COP1) with nominative subject; location expressed in a locative or oblique case adjunct for location, instrumental/comitative adjunct for companions and dative for possessors
		Identification	The subject is nominative-marked, the unique referent with which the subject referent is equated is followed by a non-verbal copula -VV-t (COP3).
		Middle	Morphological derivation, two allomorphemes (i) a glottal suffix realised as a glottal stop if the stem-final single consonant is a sonorant, realised as glottalisation if the stem-final single consonant is an obstruent, (ii) a suffix -aaq found after stem-final geminate consonants/clusters
		Passive	Morphological derivation, suffix: -am
		Reciprocal	Morphological derivation consisting of two morphemes, the middle morpheme -? / -aaq plus the passive morpheme

Affected-subject	Affected subject predication is an outcome of the combination of the verbs of class 3 with the exponent of the function suffix à. Affected subject predication is always intransitive.
Affirmative existential	á]nkwa NP
Equational	NP NP(predicate)
Goal	Two constructions code the predication. When class 1 verbs (inherently goal oriented verbs) the predication is marked by the vowel deletion on the verb preceding the noun phrase. The noun phrase that follows the verb represents the direct object. Third person singular object pronoun cannot occur on the verb. With class 2 verbs (non-affected subjects, non-affected objects) and class 3 (affected subject) verbs, the predication is marked by the suffix á (note the high tone) added to the verb.
Indirect object	Verb + (object pronouns) + 3p. sing. object pronoun n. With verbs other than inherently indirect object, the third person singular indirect object is marked by the sequence n + n. For inherently indirect object verbs, the third person singular is marked only by the pronoun n. The marking on the verb is required regardless of whether there is or not a nominal indirect object.
Indirect object role of the noun phrase	In the indirect object predication, marked by the third person object suffix n, a noun phrase is marked from the indirect object through the preposition g 'to'.
Locative	If the predicate is inherently locative, the locative predication is coded by juxtaposition of the predicate and the locative complement. If the predicate is not inherently locative, the locative predication is marked by the predictor á followed by the NP. The noun phrase is marked for the locative function by a variety of prepositions.
Negative existential	á]ká or á]ká] in clause-initial or clause-final position
Non-affected subject non-affected object	V2 +à NP
Object role of the noun phrase	When separated from the verb by other constituents, vowel deletion on the constituent preceding the NP indicates that the NP is the direct object. When following the verb of class 1, vowel reduction on the verb indicates that the noun phrase is the direct object.



BEJA – cush.

existential

overt subject + fi:/fe: 'be_there'

a construction which refers to the existence or presence of something/someone

± directional postposition dhai 'towards' ± ve
look for be_there in ge and V1.IRG in rx, and
look for SBJ in preceding context in rx; make
sure that there is no LOC, ABL, DAT, beside,
place in preceding context

KAMBAATA – cush.

Existential

Existential/copular verb yoo- (COP1) with nominative subject; location expressed in a locative or oblique case adjunct for location, instrumental/comitative adjunct for companions and dative for possessors

Intransitive predication which expresses that the subject exists, is located, is with someone/something, belongs to someone.

Look for COP1 in ge

WANDALA – chad.

Affirmative existential áŋkwà NP

Asserts the existence of an entity

#[mb="áŋkwà" & ge="exist"] NP

Formal queries allow the retrieval of relevant corpus data in the languages of CorTypo

Negative existential báakà or báakà in clause-initial or clause-final position

Denies existence of an entity

#[mb="báakà



BAGGARA ARABIC– sem.

Positive existential predication fí=PRO.a, agreement

Look for EXS in ge and/or COP in rx, at a distance of one "rx" (left) cell from a PROa

Negative existential predication má=fí=PRO.a

Look for NEG at a distance of one "ge" (left) cell from a EXS and/or NEG.PTCP at a distance of one "rx" (left) cell from

KABYLE – berb.

Negative existential Negative Existential has the form ulaf, preceded or followed by an NP in the absolute state, where the form ulaf is the negative existential predicate.

Negative existential denies the existence of a referent, or a situation. predicate. The referent

[ge=NEG.EXS & rx=PRED][ge=1 & rx=1] [ge=NOT(ABSV) & rx=PRO]\$ulaf + NO absolute pronoun expressed either by

IXCATEC – oto-mang.

existential and equational sí NP

Look for EXIST in ge and PRED in rx, AND . in ge and N in rx

Comparing Languages

- In CorpAfroAs
 - homogeneised labels (PFV for perfective, completed, ‘accompli’ ...)
 - L-defined in the grammatical sketches
 - comparison within a phylum/families
 - open comparison
 - which categories are « comparable » is to be decided by the end-user depending on their theoretical position
 - Formal categories (Newmeyer 2007)
 - ‘Comparative concepts’ (Haspelmath 2007)
 - etc...

Comparing Languages

- In CorTypo
 - Labels fully language-internal
 - defined for each language in the Table of Categories
 - categories' definitions can be compared
 - New functionalities in Elan-CorpA (constructions)
 - Separation between L-internal (corpus) and cross-linguistic (database) levels
 - Database providing domains, functions and forms + queries for retrieval of corpus data
 - Comparison guided through the database
 - which categories are comparable is prepared by the project members (through functional domains)
 - But other databases could be interfaced with the corpus (project open to further developments)

References

- Izre'el & Mettouchi (2015) Representation of Speech in CorpAfroAs: Transcriptional Strategies and Prosodic Units. In: Mettouchi, Amina, Martine Vanhove and Dominique Caubet (eds.), *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. Studies in Corpus Linguistics 68. John Benjamins: Amsterdam-Philadelphia. vi, 332 pp. + index (pp. 13–41).
- Frajzyngier, Zygmunt and Jan Mycielski. 1998. On some fundamental problems of mathematical linguistics. In Carlos Martin-Vide, ed., *Mathematical and computational analysis of natural language*. Amsterdam/Philadelphia: John Benjamins, 295-310.
- Frajzyngier, Z. & E. Shay (2003) Explaining language structure through systems interaction. Amsterdam/Philadelphia: Benjamins.
- Frajzyngier, Zygmunt. 2013. Non-aprioristic typology as a discovery tool. In *Functional-Historical Approaches to Explanation: In honor of Scott DeLancey*, eds. Tim Thornes, Erik Andvik, Gwen Hyslop and Joana Jansen. Amsterdam: John Benjamins, 3-25.
- Frajzyngier, Z. & A. Mettouchi. 2015. Functional domains and cross-linguistic comparability. In: Mettouchi, Amina, Martine Vanhove and Dominique Caubet (eds.), *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. Studies in Corpus Linguistics 68. John Benjamins: Amsterdam-Philadelphia. vi, 332 pp. + index (pp. 257-279).
- Mettouchi, A. G. Savà & M. Tosco. 2015. Cross-linguistic comparability in CorpAfroAs. In: Mettouchi, Amina, Martine Vanhove and Dominique Caubet (eds.), *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. Studies in Corpus Linguistics 68. John Benjamins: Amsterdam-Philadelphia. vi, 332 pp. + index (pp. 221-255).

Thank you



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