

Parsing Grammatical Tone Using FLEx

Jonas Lau

SOAS, University of London

WHAT I WILL TALK ABOUT TODAY

PROBLEM:

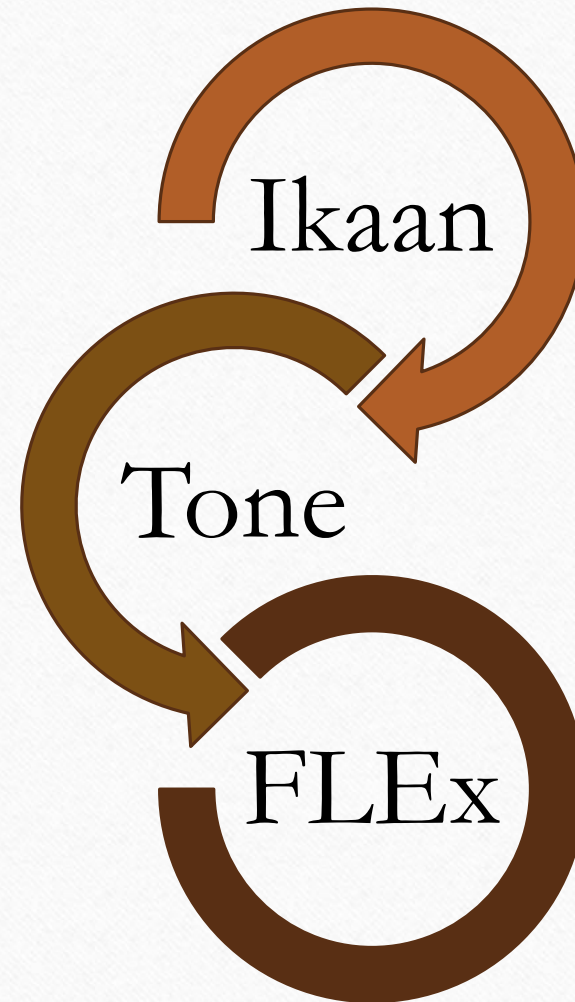
I want to use FLEx, but I am
working on a language featuring
grammatical tone

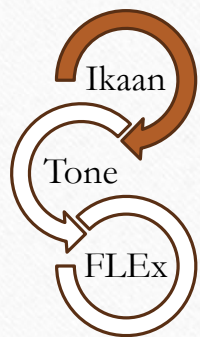


QUESTION:

How can inflectional categories
expressed by tone be glossed
automatically?

WHAT I WILL TALK
ABOUT TODAY





IKAAN

INTRODUCTION



Minority language spoken in Southern Nigeria

Around 10,000 speakers

Benue-Congo branch of Niger-Congo phylum



Tabelle Text Untertitel Lexikon Audio-Erkennen Metadaten Steuerung

Lautstärke: 100

0 50 100

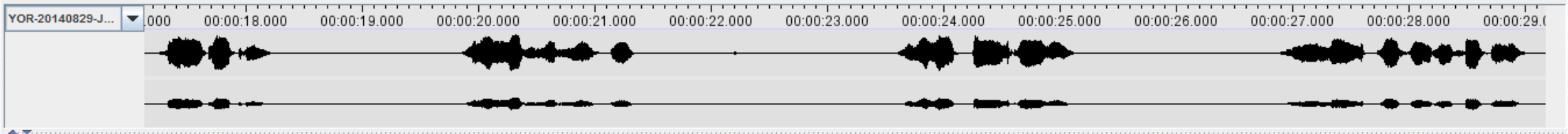
YOR-20140829-JL-001-Story001-TafaatiDayo.wav

Stumm Solo

0 25 50 75 100

Geschwindigkeit: 00:00:00.000 Auswahl: 0

PRIMARY DATA



	000	00:00:18.000	00:00:19.000	00:00:20.000	00:00:21.000	00:00:22.000	00:00:23.000	00:00:24.000	00:00:25.000	00:00:26.000	00:00:27.000	00:00:28.000	00:00:29.000			
A_morph-cf-yo [265]	i	ojo	kan	won	lo	si	inu	igbo	won	lo	pa	eran	won	ri	ef	
A_morph-gls-en [265]	t	day	one	3PL.SBJ	go	DIR	stomac	forest	3PL.SBJ	go	kill	animal	3PL.SBJ	won	see	ra
A_morph-hn-en [31]					1											
A_morph-msa-en [265]	rep	n	num	pro	v	prep	n	n	pro	v	v	n	pro	v	n	
A_morph-txt-yo [265]	i	ojo	kan	won	lo	si	inu	igbo	won	lo	pa	eran	won	ri	ef	
A_morph-type [265]	litic	stem	stem	stem	stem	clitic	stem	stem	stem	stem	stem	stem	stem	stem	stem	
A_phrase-gls-en [65]	one day,			they went into the forest.					They went to kill animals.				They saw two rabbits.			
A_phrase-segnum- [65]				15					16				17			



IKAAN

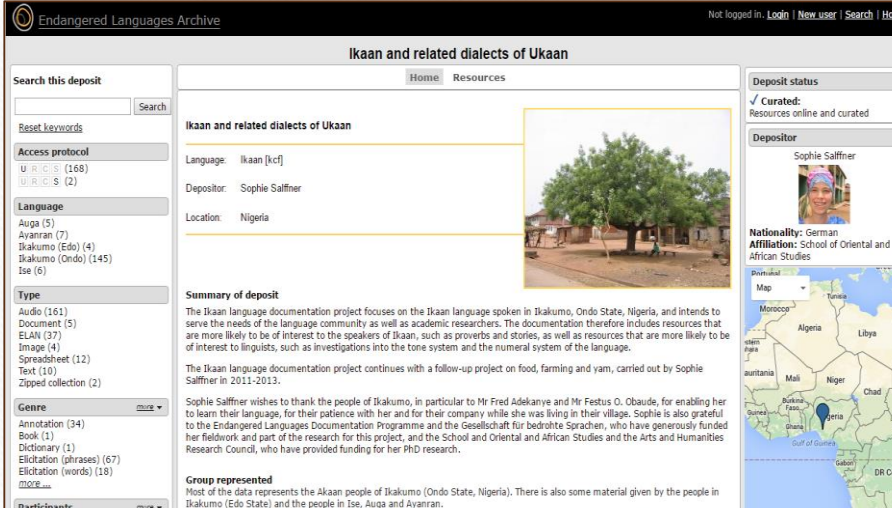
PRIMARY DATA

Deposit 0035 - **Ikaan and related dialects of Ukaan**

Sophie Salfner



THE HANS RAUSING
Endangered Languages Project
Because every last word means another lost world...



Endangered Languages Archive | Not logged in. | [Login](#) | [New user](#) | [Search](#) | [Home](#)

Ikaan and related dialects of Ukaan

Search this deposit: Search

Reset keywords

Access protocol
 U.R.C.I.S (168)
 U.R.C.S (2)

Language
 Auga (5)
 Ayanran (7)
 Ikakumo (Edo) (4)
 Ikakumo (Ondo) (145)
 Ise (6)

Type
 Audio (161)
 Document (5)
 ELAN (37)
 Image (4)
 Spreadsheet (12)
 Text (10)
 Zipped collection (2)

Genre [more](#)
 Annotation (34)
 Book (1)
 Dictionary (1)
 Elicitation (phrases) (67)
 Elicitation (words) (18)
[more](#)

Participants [more](#)

Ikaan and related dialects of Ukaan

Language: Ikaan [kcf]
 Depositor: Sophie Salfner
 Location: Nigeria

Summary of deposit
 The Ikaan language documentation project focuses on the Ikaan language spoken in Ikakumo, Ondo State, Nigeria, and intends to serve the needs of the language community as well as academic researchers. The documentation therefore includes resources that are more likely to be of interest to the speakers of Ikaan, such as proverbs and stories, as well as resources that are more likely to be of interest to linguists, such as investigations into the tone system and the numeral system of the language.
 The Ikaan language documentation project continues with a follow-up project on food, farming and yam, carried out by Sophie Salfner in 2011-2013.
 Sophie Salfner wishes to thank the people of Ikakumo, in particular to Mr Fred Adelanaye and Mr Festus O. Obaude, for enabling her to learn their language, for their patience with her and for their company while she was living in their village. Sophie is also grateful to the Endangered Languages Documentation Programme and the Gesellschaft für bedrohte Sprachen, who have generously funded her fieldwork and part of the research for this project, and the School and Oriental and African Studies and the Arts and Humanities Research Council, who have provided funding for her PhD research.

Group represented
 Most of the data represents the Akaan people of Ikakumo (Ondo State, Nigeria). There is also some material given by the people in Ikakumo (Edo State) and the people in Ise, Auga and Ayanran.

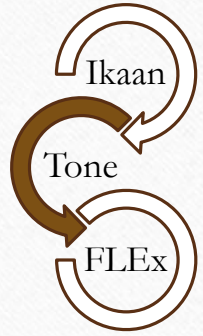
Deposit status
 ✓ Curated:
 Resources online and curated

Depositor
 Sophie Salfner

Nationality: German
Affiliation: School of Oriental and African Studies

[Map](#)





TONE

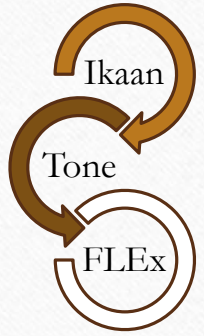
INTRODUCTION

Grammatical tone

The function of tone is not limited to distinguishing two words, but can also distinguish two grammatical categories. (Hyman 2001: 1372)

Tonal melodies

By defining tonal melodies for specific word classes, multiple correlating sequences of surface tones can be explained and grouped together. (Gussenhoven 2004: 30)



(2)	NFUT	FUT	CONT	HAB
Affirmative	dʒE LH LHL LHLL	dʒa F:H F:HH F:HHH	dʒE HD:HH HD:LHHH HD:LHHHH	dʒO F:H F:HH F:HHH
NEG Class 1	dʒE R:DH R:DHH R:DHHH		dʒO R:H R:HL R:HLL	dʒO LH LHL LHLL
NEG Class 2	dʒE R:H R:HL R:HLL		dʒO R:DH R:DHHH R:DHHH	dʒO LH LHH LHHH



(2)

Affirmative

NEG Class 1

NEG Class 2

NFUT

dʒE LH
LHL
LHLL

dʒE R:DH
R:DHH
R:DHHH

dʒE R:H
R:HL
R:HLL

FUT

dʒa F:H
F:HH
F:HHH

CONT

dʒE HD:HH
HD:LHHH
HD:LHHHH

dʒO R:H
R:HL
R:HLL

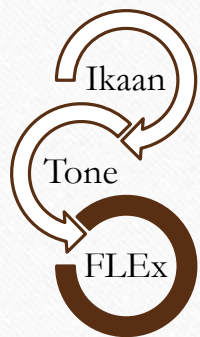
dʒO R:DH
R:DHHH
R:DHHH

HAB

dʒO F:H
F:HH
F:HHH

dʒO LH
LHL
LHLL

dʒO LH
LHH
LHHH



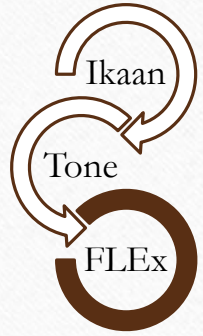
Before FLEx

TOOLBOX

McGill (2007):

- Proposes a method for Toolbox to parse grammatical tone.
- Tones are parsed as infixes and can be read together as one morpheme.
- Not applicable to FLEx

(3) \tx	ùdúkwà		
\mb	ù-	dukwa-	H L
\ge	3SG-	go	RLS
\ft	'He went'		



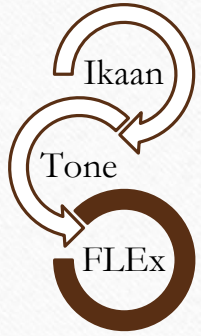
FLEx

XAmple

- Default parser
- Linear parsing: left to right

HermitCrab

- “Item and process approach” (Black 2014)
- Outside to inside
- Allows Phonological Rules and Affix Process Rules (APRs)



FLE_x

Affix Process Rules (APRs)

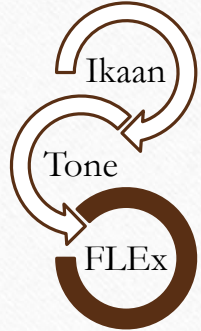
(4)

Pattern				Result
[C]	[V]	X	⇒	
1	2	3		1 2 + 1 2 3

APRs model processes to form affixes

Advantages:

- Material of the stem can be reused (reduplication)
- Different affixation strategies can be combined
- Tonal melodies can be modelled as a combination of infixes carrying one gloss



RHL₋₁ v : NFUT NFUT.NEG

☑ **Allomorphs**

[Insert Allomorph](#)

Affix Allomorph (Process: Ika)

Morph Type

prefix

Affix Process Rule

Input	[C]	[V]	[C]	[V]	[C]	[V]	X	
Index	1	2	3	4	5	6	7	

⇒ Result **R:HLL**
 ~ : 1 2 3 4 5 6 7

Affix Allomorph (Process: Ika)

Morph Type

prefix

Affix Process Rule

Input	[C]	[V]	[C]	[V]	X	
Index	1	2	3	4	5	

⇒ Result **R:HL**
 ~ : 1 2 3 4 5

Affix Allomorph (Process: Ika)

Morph Type

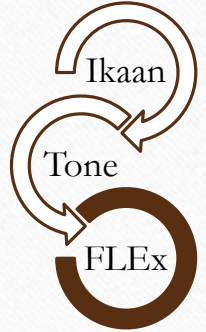
prefix

Affix Process Rule

Input	[C]	[V]	X	
Index	1	2	3	

⇒ Result **R:H**
 ~ : 1 2 3

(5)



Affix Allomorph (Process: Ika)

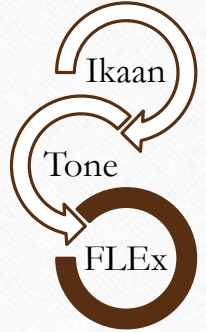
Morph Type **prefix**

Affix Process Rule

Input	[C]	[V]	[C]	[V]	X	⇒ Result ~:1 2 3 4 5
Index	1	2	3	4	5	

R:HL

(6) dzě:šénòg
 dz e ˘ : š e ´ n o ` g



Affix Allomorph (Process: Ika)

Morph Type

Affix Process Rule

prefix	
Input	[C] [V] [C] [V] X
Index	1 2 3 4 5

Result

~:1 2 3 4 5

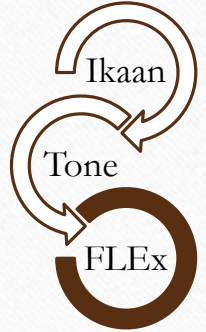
R:HL

(6) dzě:šénòg

d z e ˘ : š e ´ n o ` g

d3E-

1SG



Affix Allomorph (Process: Ika)

Morph Type

Affix Process Rule

prefix	
Input	[C] [V] [C] [V] X
Index	1 2 3 4 5

Result

~:1 2 3 4 5

R:HL

(6) dzě:šénòg

d z e ǎ : š e ´ n o `

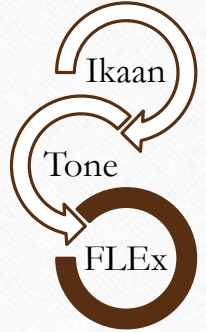
g

d3E-

-g

1SG

NEG



Affix Allomorph (Process: Ika)

Morph Type	prefix						
Affix Process Rule	Input	[C]	[V]	[C]	[V]	X	Result ~:1 2 3 4 5
	Index	1	2	3	4	5	

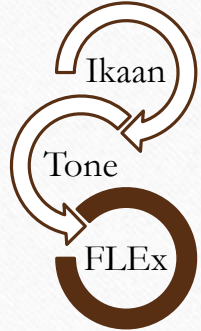
R:HL

(6) dzě:šénòg

d z e ˘ ˘ ˘ š e n o g

d3E- R:HL- -g

1SG NFUT.NEG NEG



Affix Allomorph (Process: Ika)

Morph Type

Affix Process Rule

prefix

Input	[C]	[V]	[C]	[V]	X
Index	1	2	3	4	5



Result

~:1 2 3 4 5

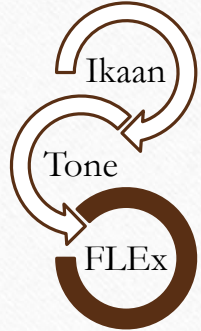
R:HL

(6) dzě:šénòg

d z e ˘ ˘ ˘ š e n o g

dzE- R:HL- šeno -g

1SG NFUT.NEG play NEG



Affix Allomorph (Process: Ika)

Morph Type

Affix Process Rule

prefix

Input	[C]	[V]	[C]	[V]	X	
Index	1	2	3	4	5	



Result

~:1 2 3 4 5

R:HL

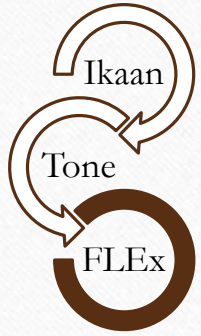
(6) dzě:šénòg

d z e ˘ ˘ ˘ š e n o g

dzE- R:HL- šeno -g

1SG NFUT.NEG play NEG

‘I did not play’



HOW TO

- Plan representation of tone in your transcriptions/orthography
 - Form paradigms
- Group several paradigms together by forming templates to regulate cooccurrences
- Create APRs with allomorphs for words of different syllable structures
- Pay attention to the order of the rules (longer affixes come first!)
 - Test with regular paradigms

CONCLUSION

Best Practice Guidelines

Four Best Practice Guidelines

- Try to keep a BALANCE between descriptive accuracy and workarounds
- Pay attention to the COMPLEXITY of the language structure, but try to minimize COMPLEXITY when modelling the structure for a parser
- Try to avoid UNDERSPECIFICATION
- Pay attention to the ORDER of rules

Literature

- Black, H. Andrew. 2014. A Conceptual Introduction to Morphological Parsing for Stage 1 of the FieldWorks Language Explorer. <http://fieldworks.sil.org/wp-content/ConceptualIntroduction.htm> (5 June, 2016).
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. (Research Surveys in Linguistics). Cambridge, New York: Cambridge University Press.
- Hyman, Larry M. 2001. Tone Systems. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language Typology and Language Universals: An International Handbook*. Vol. 2, 1689–1699. Berlin/Boston: Mouton de Gruyter. doi:10.1017/CBO9781107415324.004.
- Lau, Jonas. 2016. Parsing grammatical tone using FLEx. A case study on Ikaan negation (MA Thesis). SOAS, University of London.
- McGill, Stuart. 2007. Documenting grammatical tone using Toolbox: an evaluation of Buseman's interlinearisation technique. In Peter K Austin, Oliver Bond & David Nathan (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory*, 177–186. London: SOAS.
- Salffner, Sophie. 2010a. Ikaan and related dialects of Ukaan: an archive of language and cultural material from the Akaan people of Ikakumo (Ondo State, Nigeria). *SOAS, Endangered Languages Archive*. <http://elar.soas.ac.uk/deposit/0034> (26 July, 2016).
- Salffner, Sophie. 2010b. Tone in the phonology, lexicon and grammar of Ikaan. PhD Thesis. SOAS, University of London.
- Salffner, Sophie. 2012. Tense, aspect and manner encoding in Ikaan beyond verbal inflection. *Advances in Minority Language Research in Nigeria*.