

Using TypeCraft and Annotation Pro for multilayer annotation and analysis of Tense and Aspect in Krio

Dorothee Beermann, Katarzyna Klessa,
Beatrice Owusua Nyampong



UNIwersYTET
IM. ADAMA MICKIEWICZA
W POZNANIU

7th Language & Technology Conference November 27-29, 2015, Poznań, Poland

Overview

- ❑ the language
- ❑ the problem
- ❑ data analysis tools
- ❑ the data
- ❑ morpho-syntactic analysis
- ❑ phonetic analysis (alignment, mining, time-alignment)
- ❑ summary of findings and evaluation
- ❑ conclusion



Krio (ISO 639-3 kri)

Krio is an English-based Atlantic creole widely spoken in Sierra Leone, and other West African countries.

Next to English, the Niger-Congo languages of West Africa, but also Hausa, are *lexifier* languages for Krio.

Krio is an SVO language with a left branching NP-structure and an analytic *Tense-Aspect* system

While English is a stress language, scholars mostly agree that Krio is a tone language (Berry 1959, Mona Conference on Pidgins and Creoles 1968, Finney 2004).

The problem

Our question is whether Krio signals grammatical contrasts by pitch differences.

In the West African tone languages, tone may be lexical, so that words with different tones express different meanings, but tone may also carry grammatical information such as Tense and Aspect, and we suspect that this is also true for Krio.

The focus of our study is the Krio verb GO as pre-verb and as main verb. The multi-functionality of GO is not an isolated phenomenon but shared by other Krio verbs. Together they form a sophisticated system of analytic verbal constructions.



The
problem

Some of Krio's analytic Tense-Aspect constructions

expression	tense/aspect	specification
bin + V	tense	past
bin dɔ̃ n + V	tense/ aspect	past/perfect ??
de + V	aspect	progressive
dɔ̃ n + V	tense/aspect	perfect ??
dɔ̃ n de + V progressive	tense/aspect perfect	perfect ?? progressive
V + dɔ̃ n	aspect	completive
go + V	tense	future
go + V	aspect	inceptive

The
problem

Data analysis tools

We used Annotation Pro (Klessa, 2015) for the phonetic and TypeCraft (Beermann and Mihaylov, 2014) for the morpho-syntactic annotation of our material. Both tools facilitate multilayer annotation and annotation mining.

In TypeCraft sentence level linguistic annotation is linked to text. This made it easier to analyse the different senses of GO.

Annotation Pro was used for the investigation of perception-based and phonetic-acoustic analysis of a quasi-spontaneous narrative by a male native speaker of Krio.

The data

Our annotated TypeCraft Krio corpus consist of 8355 words in 965 phrases. It is an opportunistic corpus, consisting of transcribed short narrations, school book texts, and linguistic collections (Nyampong 2015).

We found 236 instances of go in our corpus:

78 (33%)

Vpre

59 (25%)

V

61

V1

39

V2

Total 239

For the present study we analysed the 51 instances of GO in the narration:

Noto ɔltin we fain na fain

«*Not all that glitters is gold.*»

typecraft.org >Portal of Languages>Krio>

The data

Virtual workspace for our morpho-syntactic analysis

TC Editor - Google Chrome

typecraft.org/tc2/tceditor.html#2856,54477

Text > Phrase > Theme >

Save Share: Private Publish Template New phrase Delete phrase View phrase list View discourse senses

B I U AaBbCc Paragraph AaBb Heading 1 AaBb Heading 2 AaBb Heading 3 AaBb Heading 4 AaBbCc Pre Remove formatting

Language: Krio Change

Title: Nɔto ɔltin we fain na fain (124 phrases, 43 annotation OI)

Title translation: Not all that glitters is gold.

Content description: A folktale about a girl who married a stranger.

Wan de ya, na wan titi bin de, i de na wan fa we vilej wit in mama en papa en in smɔl brɔda. Bot i bin de na po po famili so i bin fotunet se na in den sen na Fritɔŋ fɔ go skul. We i dɔn di skul nain i get fɔ ritɔn bak na di vilej. So nain in mama tel am se 'Eee Bindu, ɔl wi mɔni we wi dɔn get nain wi dɔn put pan dis yu skul. I se 'So duya trai mared we wi sef go ebul fɔ get smɔl tin. Nain di pikin tel am se, 'Ee mama, mi a nɔ si no man na dis vilej oo'. So Bindu tok se i nɔ si no man na dis vilej ya we go fit am. Nain in mama se 'Wel aw yu wan mek wi du naw'. I se 'di wan der Bot ya so if yu mared Nain Bindu se, 'mama.

10. If yu se pas yu ...

Save FTrans 1 FTrans 2 CParam Base Meaning Gloss POS

Phrase: If yu se pas yu go bak na Fritɔŋ fɔ go fen man, wi nɔ no bot den.

Free translation 1: If you want to go back to Freetown to look for a husband we do not know about them.

Valence: Change

Word:	if	yu	se	pas	yu	go	bak	na	Fritɔŋ	fɔ	go	fen	mar	,	Wi
Morph:	if	yu	se	pas	yu	go	bak	na	fritɔŋ	fɔ	go	fen	mar	,	wi
Citation Form:	if	yu	se	pas	yu	go	bak	na	fritɔŋ	fɔ	go	fen	mar	,	wi
Meaning:			say			go			Freetown		go	find	mar		
Gloss tags:			2SG		CMPR	2SG					INCEP				1PL
POS:	CONJ	PN	V	CON	PN	V	AD	PREI	Nf	CON	Vpre	V2	N	PUI	PN

Add discourse sense

So go on go on, wan fain man. I de waka c dɔn si di man we a go de spot, I de waka, i kagbɔ'. I se 'duya, mi kin gi mi taim fɔ lisin t

Nain di debul se, 'ma we mi sef go lek'. So kam, den kam gi am Bindu, 'Bindu na yu s nɔ de chen a lek dis fɔ le den kam du di m dem. I briŋ wan bɔks

So Bindu naw, we de na baba we bin est e

analysis

Morpho-syntactic analysis (Nyampong 2015)

First cycle of
morpho-syntactic
analysis

Tone value suggested
in the literature

GO, Vpre, TENSE= FUTURE

Tone = LT

GO, V, TENSE= PAST, PRED 'walk'

Tone = HT

GO, V TENSE= unmarked , PRED 'walk'

Tone unknown if any

GO, Vpre, ASPECT=INCEPTIVE

Tone unknown if any

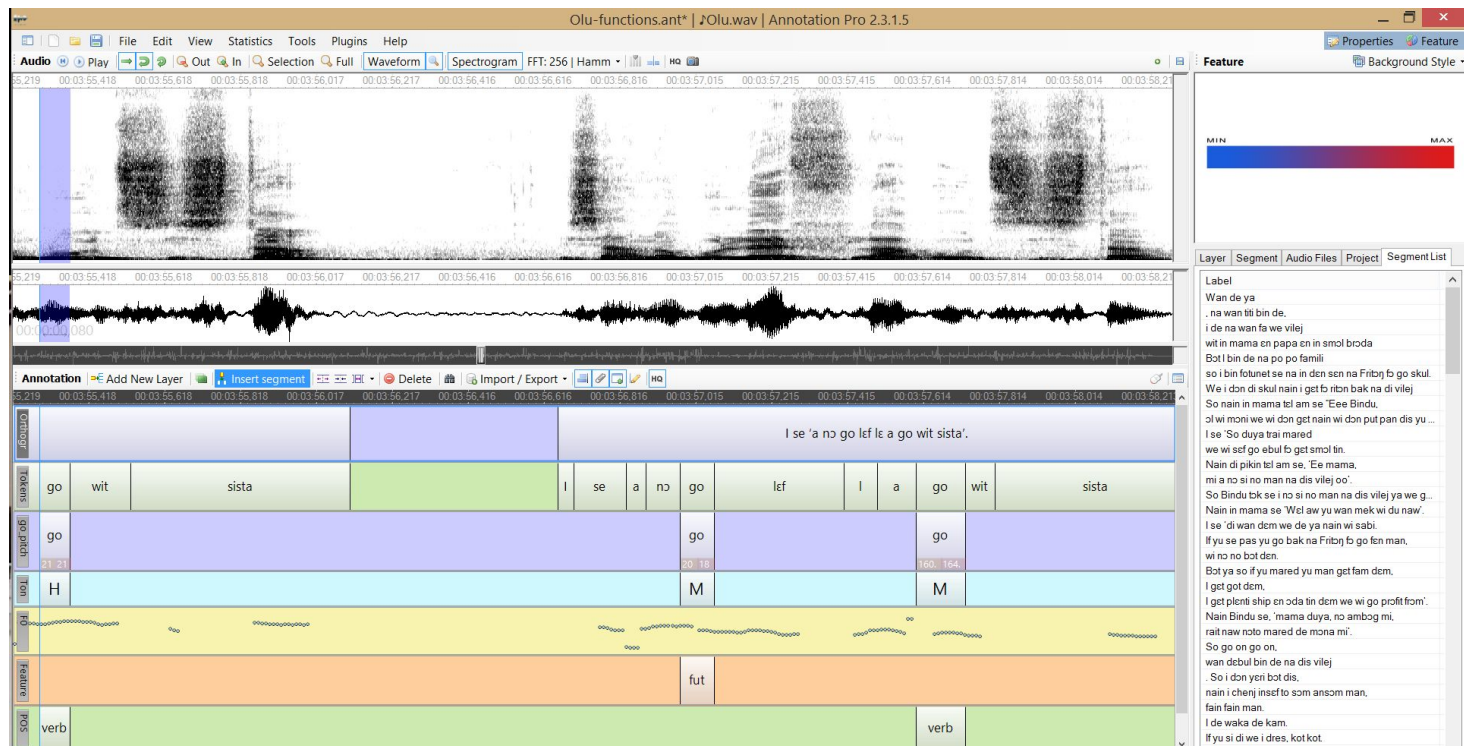
As we annotate more of our data we discover new meanings elucidating the use of GO and other Krio verbs that do double duty of main and pre-verbs

for example:

«lisin to mi a go gladi.» GO, V, PRED 'become'

analysis

Time-alignment and annotation mining



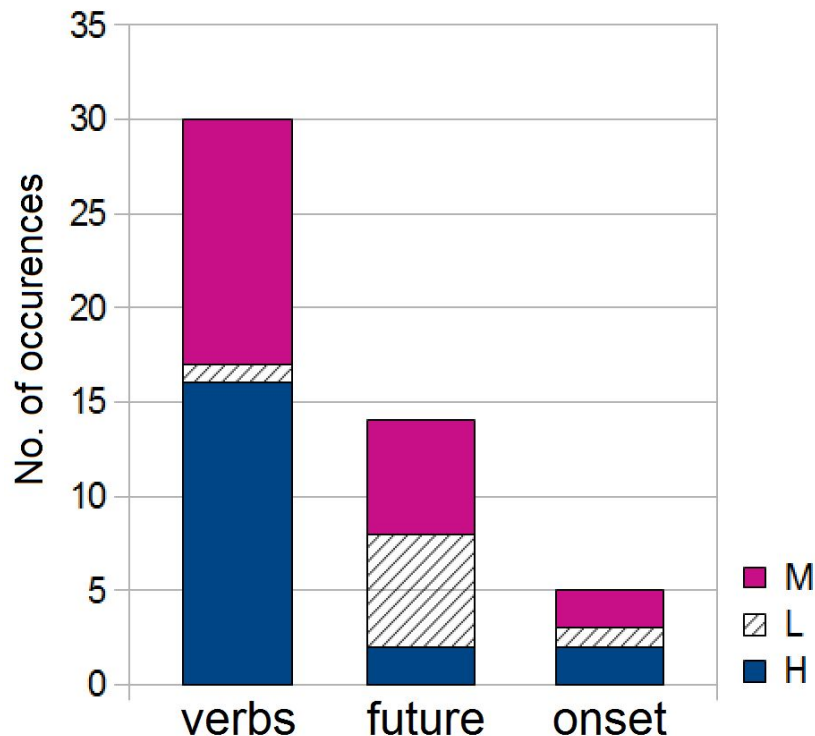
alignment

Time-alignment and annotation mining

- ❑ Import of orthographic annotations from TypeCraft (via TXT)
- ❑ Time alignment on the level of phrases
- ❑ Automatic segmentation into words with SPPAS (Bigi, 2015) English model
- ❑ Manual corrections
- ❑ Perception-based labelling of tone level (L M H)
- ❑ Automatic extraction of f0, and intensity with Praat (Boersma, Weenink, 2015) -> imported back to Annotation Pro
- ❑ Automatic duration extraction with Annotation Pro

mining

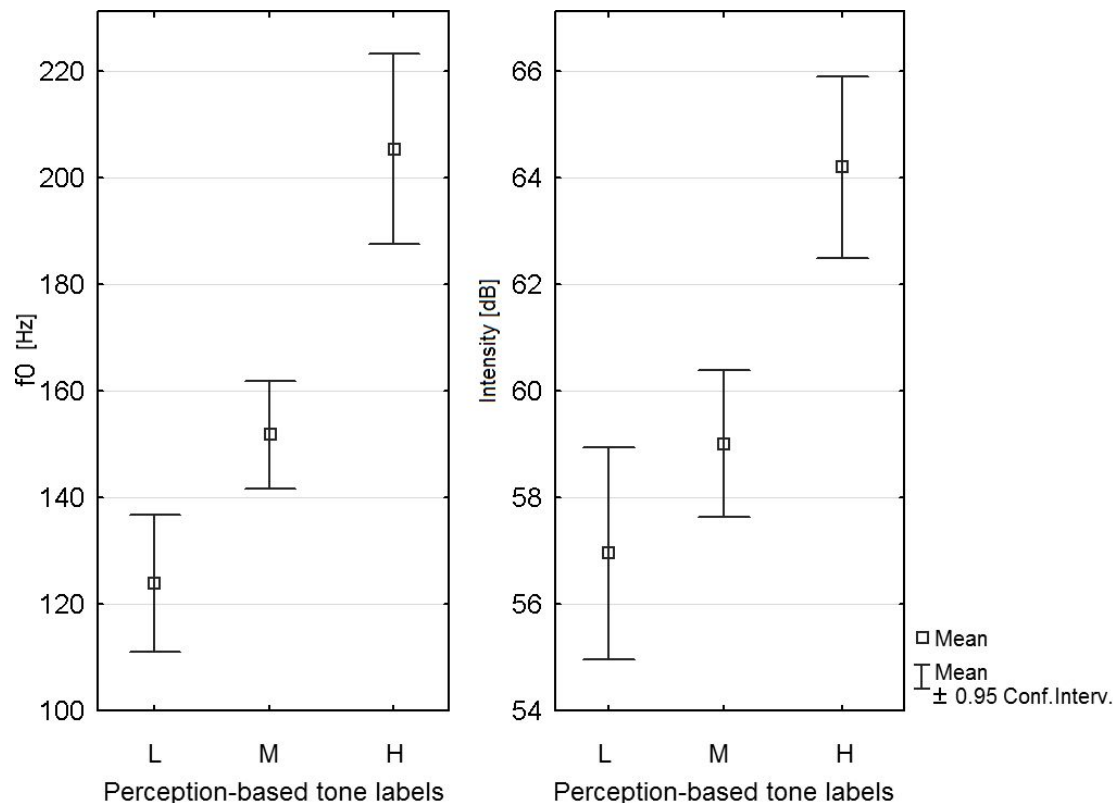
Perception-based judgements of tone



- Native speaker female labeller
- A three-level tone notation: L (Low tone), M (Mid tone), H (High tone);
- Verbs - mainly H, M;
- Future markers - mainly L, M

Time-alignment

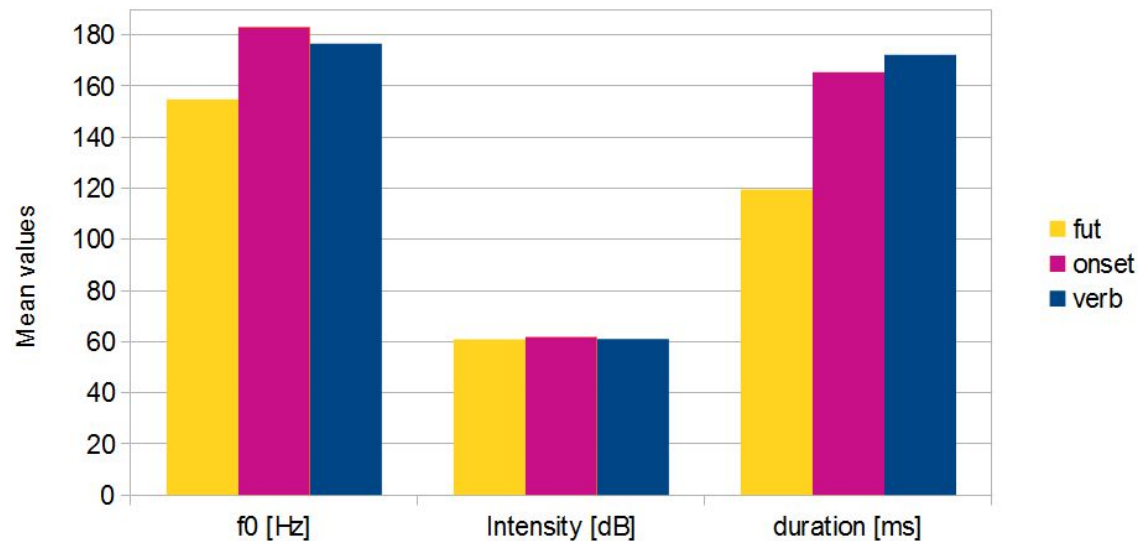
Acoustic-phonetic measures vs. perception-based labels



- Perception-based tone labels positively correlated with both f_0 and intensity
- Durational variability insignificant between the levels

Time-alignment

Acoustic-phonetic measures vs. functions of 'GO'



The relationships are not obvious but certain tendencies can be observed:

- ☐ Mean durations longer in verbs
- ☐ *f0* lower in future markers
- ☐ intensity - no systematic differences

Time-alignment

Forms & functions of 'GO': summary of findings

Feature	Earlier studies	Perceived	Measured
go, V, non-past	no prediction	V class - tendency medium-high	rising contour, 6 verbs with a rising-falling one
go, V, past	high tone		
go, Vpre, future	low tone	tendency medium-low	10 cases with a falling contour, lower <i>f0</i> means
go, Vpre, onset-event	no predication	no tendency	no significant differences to main verb

Table 1. Summary of findings with regards to form and function of Krio GO.

evaluation

Conclusion - analysis

GO	V	TENSE= PAST NON-PAST SENSE1: 'go','walk', 'leave' ... SENSE2: 'become' SENSE3: GOAL.LOCATION
GO +V	Vpre	Aspect=INCEPTIVE
GO +V	Vpre	TENSE= FUTURE

- expressed by tone (either the past, or the distinction between past and non-past)

FURTHER WORK:

- ❑ annotation of already collected data, and improvement of the existing annotations
- ❑ extension of the present approach to the other tense-aspect constructions

conclusion

CONCLUSION - methodology

DONE:

- ❑ Efficiently combining information from independent software tools to analyse lesser-resourced language data
- ❑ Acoustic and perception-based support for tone labelling
 - ❑ annotation based solely on textual data impossible due to uncertainty and lack of common standards
- ❑ Application of an English acoustic model for automatic segmentation of Krio: sufficient for small data but for larger corpora - tuning needed

FURTHER WORK:

- ❑ Towards better interoperability between TypeCraft & Annotation Pro - implementation of more sophisticated import/export options
- ❑ Tone labelling - usage of automatized procedures, e.g. by implementing a new Annotation Pro plugin (cf. MOMEL, Prosogram or other).

conclusion

Thank you

Contact:

dorothee.beermann@ntnu.no
klessa@amu.edu.pl
nyampong.beatrice@gmail.com

NTNU, Trondheim, Norway
Adam Mickiewicz University, Poznań, Poland