

Describing Morphologically-rich Languages using Metagrammars: a Look at Verbs in Ikota

Denys Duchier¹, Brunelle Magnana Ekoukou², Yannick Parmentier¹, Simon Petitjean¹, Emmanuel Schang²

¹ LIFO, Université d'Orléans - 6, rue Léonard de Vinci 45067 Orléans Cedex 2 – France
{denys.duchier,yannick.parmentier,simon.petitjean}@univ-orleans.fr

² LLL, Université d'Orléans - 10, rue de Tours 45067 Orléans Cedex 2 – France
{brunelle.magnana-ekoukou,emmanuel.schang}@univ-orleans.fr

Purpose

- Provide a formal description of the morphology of verbs in Ikota
- Automatically derive from this description a lexicon of inflected forms

- Lexicalized wide-coverage tree-grammars for natural languages are very large and extremely resource intensive to develop and maintain
- They can be automatically produced by software from a highly modular formal description called a metagrammar, introduced by [1]
- Metagrammars are much easier to develop and to maintain
- We propose to adopt a similar strategy to capture morphological generalizations over verbs in Ikota

Ikota language

Ikota is a Bantu language of Gabon and the Democratic Republic of Congo. It is threatened with extinction in Gabon, mainly because of its abandon for French. It shares many grammatical features with the Bantu languages. Ikota is a tonal language with two registers (High and Low). It has ten noun classes and has a widespread agreement in the NP.

Table: Ikota's noun classes

Noun class	prefix	allomorphs
CL 1	mò-, Ø-	mw-, ñ-
CL 2	bà-	b-
CL 3	mò-, Ø-	mw-, ñ-
CL 4	mè-	
CL 5	ì-, ɕ-	dy-
CL 6	mà-	m-
CL 7	è-	
CL 8	bè-	
CL 9	Ø-	
CL 14	ò-, bò-	bw-

Verbs in Ikota

Verbs consist of a lexical root (VR) and several affixes distributed on each side of the VR.

Table: Verb formation

Subj-	Tense-	VR	-(Aspect)	-Active	-(Proximal)
-------	--------	----	-----------	---------	-------------

Table: Verbal forms of bòɕákà "to eat"

Subj.	Tense	VR	Aspect	Active	Prox.	Value
m-	à-	ɕ		-á		present
m-	à-	ɕ		-á	-ná	past, yesterday
m-	à-	ɕ		-á	-sá	distant past
m-	é-	ɕ		-á		recent past
m-	é-	ɕ	-àk	-à		medium future
m-	é-	ɕ	-àk	-à	-ná	future, tomorrow
m-	é-	ɕ	-àk	-à	-sá	distant future
m-	ábí-	ɕ	-àk	-à		imminent future

The infinitival suffix AKA at the lexical level can be realized at the surface level by ákà, éɕè, ókò, depending on the verb class.

References

- [1] Marie Candito, *A Principle-Based Hierarchical Representation of LTAGs*, Proceedings of the 16th International Conference on Computational Linguistics (COLING'96)
- [2] Magnana Ekoukou, Brunelle, *Morphologie nominale de l'ikota (B25): inventaire des classes nominales* Mémoire de Master 2, Université d'Orléans, 2010
- [3] Piron, Pascale, *Éléments de description du kota, langue bantoue du Gabon, mémoire de licence spéciale africaine*, Université Libre de Bruxelles, 1990

eXtensible MetaGrammar

XMG is normally used to describe lexicalized tree grammars. In other words, an XMG specification is a declarative description of the tree-structures composing a grammar. This description relies on four main concepts:

- **abstraction**: the ability to associate a content with a name
- **contribution**: the ability to accumulate information in any level of linguistic description
- **conjunction**: the ability to combine pieces of information
- **disjunction**: the ability to non-deterministically select pieces of information

Formally, one can define an XMG specification as follows:

$Rule := Name \rightarrow Content$
 $Content := Contribution \mid Name \mid Content \vee Content \mid Content \wedge Content$

Metagrammar of Ikota verbal morphology

<i>Subj</i>	\rightarrow	$\begin{matrix} 1 \leftarrow m \\ p = 1 \\ n = sg \end{matrix} \vee \begin{matrix} 1 \leftarrow \dot{o} \\ p = 2 \\ n = sg \end{matrix} \vee \dots$
<i>Tense</i>	\rightarrow	$\begin{matrix} 2 \leftarrow \acute{e} \\ tense = past \\ proxi = near \end{matrix} \vee \begin{matrix} 2 \leftarrow \acute{e} \\ tense = future \end{matrix} \vee \begin{matrix} 2 \leftarrow \grave{a} \\ tense = present \end{matrix} \vee \begin{matrix} 2 \leftarrow \grave{a} \\ tense = past \\ proxi = \neg near \end{matrix} \vee \begin{matrix} 2 \leftarrow \acute{a}b\acute{i} \\ tense = future \\ proxi = imminent \end{matrix}$
<i>Active</i>	\rightarrow	$\begin{matrix} 5 \leftarrow \grave{A} \\ active = + \\ prog = - \end{matrix} \vee \begin{matrix} 5 \leftarrow \acute{A} \\ active = + \\ prog = + \end{matrix} \vee \begin{matrix} 4 \leftarrow \acute{e}bw\grave{e} \\ active = - \end{matrix}$
<i>Aspect</i>	\rightarrow	$\begin{matrix} 4 \leftarrow \grave{A}K \\ tense = future \\ prog = - \end{matrix} \vee \begin{matrix} tense = \neg future \\ prog = + \end{matrix}$
<i>Proximal</i>	\rightarrow	$\begin{matrix} 6 \leftarrow n\acute{A} \\ proxi = day \end{matrix} \vee \begin{matrix} 6 \leftarrow s\acute{A} \\ proxi = far \end{matrix} \vee \begin{matrix} proxi = none \vee near \end{matrix} \vee \begin{matrix} proxi = imminent \\ tense = future \end{matrix}$
<i>To-Eat</i>	\rightarrow	$\begin{matrix} 3 \leftarrow \mathfrak{c} \\ vclass = g1 \end{matrix}$
<i>To-Give</i>	\rightarrow	$\begin{matrix} 3 \leftarrow w \\ vclass = g2 \end{matrix}$
<i>VR</i>	\rightarrow	$To-Eat \vee To-Give$
<i>Verb</i>	\rightarrow	$Subj \wedge Tense \wedge VR \wedge Aspect \wedge Active \wedge Proximal$

Figure: A successful derivation for : ò+é+ɕ+ÀK+Ā+nĀ → ò+é+ɕ+àk+à+ná → óɕàkàná

<i>Verb</i>	\rightarrow	$Subj \wedge Tense \wedge VR \wedge Aspect \wedge Active \wedge Proximal$
\rightarrow		$\begin{matrix} 1 \leftarrow \dot{o} \\ p = 2 \\ n = sg \end{matrix} \wedge \begin{matrix} 2 \leftarrow \acute{e} \\ tense = future \end{matrix} \wedge \begin{matrix} 3 \leftarrow \mathfrak{c} \\ vclass = g1 \end{matrix} \wedge \begin{matrix} 4 \leftarrow \grave{A}K \\ tense = future \\ prog = - \end{matrix} \wedge \begin{matrix} 5 \leftarrow \grave{A} \\ active = + \\ prog = - \end{matrix} \wedge \begin{matrix} 6 \leftarrow n\acute{A} \\ proxi = day \end{matrix}$
\rightarrow		$\begin{matrix} 1 \leftarrow \dot{o} & 2 \leftarrow \acute{e} & 3 \leftarrow \mathfrak{c} & 4 \leftarrow \grave{A}K & 5 \leftarrow \grave{A} & 6 \leftarrow n\acute{A} \\ p = 2 & prog = - & tense = future & vclass = g1 & active = + & proxi = day \\ n = sg & active = + & proxi = day \end{matrix}$

Figure: A failed derivation: clashes on tense and on prog

<i>Verb</i>	\rightarrow	$Subj \wedge Tense \wedge VR \wedge Aspect \wedge Active \wedge Proximal$
\rightarrow		$\begin{matrix} 1 \leftarrow \dot{o} \\ p = 2 \\ n = sg \end{matrix} \wedge \begin{matrix} 2 \leftarrow \acute{e} \\ tense = future \end{matrix} \wedge \begin{matrix} 3 \leftarrow \mathfrak{c} \\ vclass = g1 \end{matrix} \wedge \begin{matrix} tense = \neg future \\ prog = + \end{matrix} \wedge \begin{matrix} 5 \leftarrow \grave{A} \\ active = + \\ prog = - \end{matrix} \wedge \begin{matrix} 6 \leftarrow n\acute{A} \\ proxi = day \end{matrix}$
\rightarrow		failure!

Conclusion

- We proposed a formal, albeit preliminary, declarative description of verbal morphology in Ikota, an arguably minority African language
- From this formal description, using XMG, we are able to automatically produce a lexicon of fully inflected verb forms with morphosyntactic features.
- Methodology : XMG helps to express ideas, test them quickly, and then validate the results against the available data
- Using the same tool, we will be able to also describe the syntax of the language using e.g. tree-adjoining grammars (the topic of an ongoing PhD thesis)