

PFM : pour une implémentation de la morphologie de l'ikota dans XMG

Brunelle Magnana Ekoukou

LLL, Université d'Orléans - 10, rue de Tours 45067 Orléans Cedex 2

magnanabrunelle@yahoo.fr

Résumé. Cet article traite de la représentation formelle de la morphologie des noms et des verbes de l'ikota, langue bantoue du Gabon. J'utilise la notion de classes de position (CP) de la PMF (Paradigm Function Morphology), théorie morphologique qui s'intéresse à la flexion. Les noms et les verbes seront représentés en fonction de leurs CP. Ce procédé sera réutilisé dans le langage XMG (eXtensible MetaGrammar) afin d'implémenter la morphologie de cette langue.

Abstract. This article discusses the formal representation of the morphology of nouns and verbs of ikota, bantu language of Gabon. I use the notion of position class (CP) of PFM (Paradigm Function Morphology), morphological theory is interested in the inflection. The nouns and verbs will be represented according to their CP. This process will be re-used in the XMG language (eXtensible MetaGrammar) to implement the morphology of this language.

Mots-clés : PFM, implémentation, morphologie, ikota.

Keywords: PFM, implementation, morphology, Ikota.

1 PFM

PFM fait partie des modèles morphologiques traitant particulièrement de la flexion (Stump, 1992, 1998, 2001). C'est une théorie inférentielle et réalisationnelle qui définit le système flexionnel d'une langue comme une fonction paradigmatique mettant en jeu un lexème et ses propriétés morphosyntaxiques pour déterminer la forme flexionnelle d'un mot. La fonction paradigmatique est considérée comme une règle de réalisation. Il en existe deux types :

- La règle d'exposant spécifie les modifications apportées à la racine d'un lexème ;
- La règle de renvoi reporte la réalisation d'un ensemble de propriétés morphosyntaxiques à celle d'un autre ensemble de propriétés morphosyntaxiques. Cette règle est utilisée dans le cas des phénomènes de syncrétisme, fusion en un seul élément de plusieurs propriétés morphosyntaxiques.

$$n, X_c, t \longrightarrow f(X)$$

Figure 1: Format de base des règles de réalisation

n représente le numéro du bloc dans lequel la règle s'applique, X la racine du lexème, C la classe du lexème (sa catégorie), t les propriétés morphosyntaxiques du lexème et $f(X)$ la forme phonologique résultant de l'application de la règle. Les règles de réalisation sont organisées en blocs (Anderson, 1992; Stump, 2001; Stewart & Stump, 2007) de telle sorte que les règles qui appartiennent au même bloc sont en compétition pour la même position. Cette compétition entre blocs de règles est arbitrée par le principe de *Pānini* (Stump, 1998). Une règle de réalisation peut s'étendre sur plusieurs blocs, on parle dans ce cas de « *portmanteau* ». En plus des règles de réalisation, Stump (2001) introduit le concept de CP. Les CP sont des positions qui constituent la structure préexistante des mots ou des phrases.

2 PFM appliqué à la morphologie de l'ikota

L'ikota (B25) est une langue bantoue, parlée au Gabon et en République du Congo. Au Gabon, cette langue est menacée de disparition principalement en raison de l'influence du français (langue officielle du pays). Cette langue manifeste plusieurs traits (Piron, 1990; Magnana Ekoukou, 2010) :

- C'est une langue tonale avec deux niveaux de hauteur (haut et bas) ;
- C'est une langue à classes nominales, avec dix classes répertoriées. Les classes nominales informent sur la propriété *nombre*. La notion de genre telle qu'elle est perçue dans plusieurs langues européennes n'existe pas en ikota.

Classes nominales	Exposants
1	mò, -
2	bà
3	mò, ò
4	mè
5	ì, ð
6	mà
7	è
8	bè
9	-
14	bò, ò

Table 1: Classes nominales de l'ikota

2.1 Les noms

Le nom en ikota se compose de deux CP. La CP 1 marque l'exposant nominal. La CP 2 marque le stem :

1	-	2
Nom:	Exposant nominal	Stem

- (1) mò - tò → mòtò "homme"
 è - dó → èdó "hache"

La CP 1 peut être vide contrairement à la CP 2 :

- (2) - ndú → ndú "tambour"
 mà - ndú → màndú "tambours"

Des règles de réalisation peuvent être établies car les CP du nom marquent une information et une seule. Comme les règles ne peuvent introduire qu'un exposant à la fois, les classes qui ont une CP 1 occupée par plus d'un exposant vont être fractionnées en sous-classes. Je suis donc amenée à revoir en partie la classification traditionaliste et à diviser par exemple la classe 1 en sous-classes 1a et 1b, la classe 3 en sous-classes 3a, 3b et 3c, la classe 5 en sous-classes 5a et 5b, et enfin la classe 14 en sous-classes 14a et 14b. Pour les noms de classe 1a par exemple, on aura la règle suivante :

1- $X_N, \sigma: \{C11a\} \rightarrow m\delta X$

la règle 1 réalise la propriété {*classe 1a*} par la préfixation de *mò* à la racine d'un lexème de catégorie *N*.

2.2 Les verbes

Les verbes de l'ikota sont répartis dans trois groupes en fonction des suffixes verbaux. Dans la conjugaison, les suffixes verbaux sont à l'origine d'un phénomène d'harmonie vocalique. Les tableaux (2), (3) et (4) montrent l'analyse des CP de trois verbes conjugués à la première personne.

1	2	3	4	5	6	7	Valeur
m-	à-	ɕ			-á		présent
m-	à-	ɕ			-á	-ná	passé d'hier
m-	à-	ɕ			-á	-sá	passé lointain
m-	é-	ɕ			-à		passé récent
m-	àmò-	ɕ			-á		passé moyen
m-	é-	ɕ		-ák	-à		futur moyen
m-	é-	ɕ		-ák	-à	-ná	futur de demain
m-	é-	ɕ		-ák	-à	-sá	futur lointain
m-	ábí-	ɕ		-ák	-à		futur imminent

Table 2: Conjugaison de *bòɕákà* « manger » (groupe 1)

1	2	3	4	5	6	7	Valeur
m-	à-	w			-é		présent
m-	à-	w			-é	-né	passé d'hier
m-	à-	w			-é	-sé	passé lointain
m-	é-	w			-è		passé récent
m-	àmò-	w			-é		passé moyen
m-	é-	w		-éɥ	-è		futur moyen
m-	é-	w		-éɥ-	è	-né	futur de demain
m-	é-	w-		-éɥ	-è	-sé	futur lointain
m-	ábí-	w		-éɥ	-è		futur imminent

Table 3: Conjugaison de *bòwéɥè* « donner » (groupe 2)

1	2	3	4	5	6	7	Valeur
m-	à-	bón			-ó		présent
m-	à-	bón			-ó	-nó	passé d'hier
m-	à-	bón			-ó	-só	passé lointain
m-	é-	bón			-ò		passé récent
m-	àmò-	bón			-ó		passé moyen
m-	é-	bón		-ók	-ò		futur moyen
m-	é-	bón		-ók	-ò	-nó	futur de demain
m-	é-	bón		-ók	-ò	-só	futur lointain
m-	ábí-	bón		-ók	-ò		futur imminent

Table 4: Conjugaison de bòbónókò « choisir » (groupe 3)

De ces tableaux, il ressort que les formes verbales de l'ikota sont composées de sept CP :

- La CP 1 marque l'indice sujet ;
- La CP 2 est occupée par un exposant ayant rapport au temps ;
- La CP 3 marque le stem ;
- La CP 4 marque la voix. Cette position est vide à la voix active et pleine à la voix passive ;
- La CP 5 marque l'aspect (progressif ou non) ;
- La CP 6 est occupée par un exposant qui marque la voyelle thématique ;
- La CP 7 marque l'éloignement.

Le tableau (5) présente la structure du verbe en ikota.

Indice sujet-	Indice temporel-	Stem	-(Voix)	-(Aspect)	-Voyelle thématique	-(Éloignement)
---------------	------------------	------	---------	-----------	---------------------	----------------

Table 5: Structure du verbe

Étant donné que la CP 2 a un statut ambigu, aucune règle ne sera proposée pour les verbes. Un temps verbal sera considéré comme la concaténation simultanée de plusieurs CP.

3 PFM et l'implémentation dans le langage XMG

XMG est à la fois un langage formel et un logiciel de compilation de méta-grammaire (Crabbé *et al.*, 2012). Conçu à l'origine pour décrire les grammaires d'arbres adjoints (Duchier *et al.*, 2005; Parmentier *et al.*, 2006), XMG a déjà été utilisé pour décrire la morphologie verbale de l'ikota (Duchier *et al.*, 2012). PFM permet l'implémentation de la morphologie de l'ikota dans XMG car son concept de CP et ses règles de réalisation (en ce qui concerne les noms) peuvent être réutilisés dans XMG. La formalisation dans XMG utilise la notion de *domaine topologique* (Bech, 1955) qui consiste en une séquence linéaire de champs organisée dans des blocs¹ élémentaires. Un bloc élémentaire va fournir deux types d'informations : l'information sur la phonologie qui prend en compte la forme phonologique lexicale des items (exposant nominal et stem) et l'information sur la flexion qui prend en compte les propriétés morphosyntaxiques

¹Le bloc dans le langage XMG fait référence à une règle qui définit comment une abstraction peut être décrite

propres à chaque item. Dans le langage XMG, à un champ doit correspondre un item et un seul qui représente la forme phonologique lexicale d'un exposant.

3.1 PFM et l'implémentation des noms

En s'inspirant de la notion de CP, le nom dans le langage XMG est défini comme la concaténation de deux blocs élémentaires. Le premier bloc représente la forme phonologique de l'exposant et son trait morphosyntaxique. Le second bloc représente le stem. Le tableau (6) montre la structure du nom dans le langage XMG.

1 → Préfixe	∧	2 → RN (racine nominale)
nc		

Table 6: Structure du nom en XMG

Le trait *nc* fait référence à la classe nominale. Comme en PFM, la description dans le langage XMG va utiliser des règles permettant d'introduire les exposants en CP 1. Pour les noms de classe 1a par exemple on aura la notation suivante :

```
class prefix
{
    <morph>{
    {
    { nc=C1a; prefix <- "mò" }
    }
}
```

Cette notation veut dire que lorsque la classe est 1a, *mò* doit être préfixé à la racine nominale.

3.2 PFM et l'implémentation des verbes

Duchier *et al.* (2012) proposent d'implémenter les verbes de l'ikota en s'appuyant sur le concept de CP. Aux sept CP relevées dans le tableau (5) va correspondre sept blocs élémentaires dans XMG :

PFM	XMG
1. Indice sujet	Subject
2. Indice temporel	Tense
3. Stem	RV (racine verbale)
4. Voix	Voice
5. Aspect	Aspect
6. Voyelle thématique	Theme
7. Éloignement	Proximal

Table 7: Correspondance de la structure du verbe en PFM et XMG

Le tableau (8) montre la description de *méçákàná* « je mangerai (futur de demain) » dans le langage XMG. Cette forme verbale résulte de la concaténation simultanée de sept CP :

1 ← m	2 ← é	3 ← ç	4 ← nul	5 ← Ák	6 ← À	7 ← nÁ
p = 1 n = sg	tense = futur	g1	active = +	tense = futur prog = -	theme = g1	proxi = day

Table 8: Formalisation de *méçákàná* « je mangerai (futur de demain) »

4 Conclusion et perspectives

En utilisant PFM, je me suis intéressée dans cet article à l'établissement de l'ordre des positions dans un mot flexionnel (verbal ou nominal). Ce procédé a permis d'implémenter la morphologie. Toutes les catégories grammaticales (adjectifs, connectifs, déterminants (démonstratifs et possessifs) etc.) de l'ikota sont susceptibles d'être décrites dans ce formalisme car, comme les noms et les verbes, elles peuvent être représentées sous forme de CP. L'adjectif *ùénè* « grand » par exemple se compose de deux CP : la CP 1 marque l'exposant nominal et la CP 2 marque le stem. La description de la conjugaison de trois verbes à l'actif et au passif, en incluant la négation, permet d'obtenir environ 600 formes verbales fléchies. celles-ci peuvent être exportées au format XML pour une éventuelle réutilisation. Pour une langue peu dotée comme l'ikota, il serait intéressant de pouvoir produire de manière automatique des ressources lexicales monolingues ou bilingues par exemple comme cela est fait en LMF (Lexical Markup Framework).

Références

- ANDERSON S. (1992). *A-morphous morphology*, volume 62. Cambridge University Press.
- BECH G. (1955). Studien über das deutsche verbum infinitum. *Danske Videnskabernes Selskab*.
- CRABBÉ B. & AL. (2003). Une plateforme de conception et d'exploitation de grammaire d'arbres adjoints lexicalisés. In *Actes de la conférence TALN'2003*.
- CRABBÉ B., DUCHIER D., GARDENT C., LE ROUX J. & PARMENTIER Y. (2012). Xmg : extensible metagrammar. *Computational Linguistics*, p. 1–39.
- CRABBÉ, B. ET DUCHIER D. (2005). Metagrammar redux. In *Constraint Solving and Language Processing*, p. 32–47. Springer.
- DUCHIER D., LE ROUX J., PARMENTIER Y. & NANCY H. (2005). Xmg : Un compilateur de méta-grammaires extensible. *Actes de TALN 05*.
- DUCHIER D., MAGNANA EKOUKOU B., PARMENTIER Y., PETITJEAN S., SCHANG E. *et al.* (2012). Décrire la morphologie des verbes en ikota au moyen d'une métagrammaire. *JEP-TALN-RECITAL 2012*, p.97.
- MAGNANA EKOUKOU B. (2010). *Morphologie nominale de l'ikota : inventaire des préfixes de classes nominales*. Mémoire de master 2, université d'Orléans.

PARMENTIER Y., LE ROUX J. & CRABBÉ B. (2006). Xmg: an expressive formalism for describing tree-based grammars. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics : posters & demonstrations*, p. 103–106: Association for Computational Linguistics.

PIRON P. (1990). *Éléments de description du kota, langue bantoue du Gabon*. Mémoire de licence spéciale africaine, université libre de Bruxelles.

STEWART T. & STUMP G. (2007). Paradigm function morphology and the morphology-syntax interface. In *The Oxford handbook of linguistic interfaces*: Citeseer.

STUMP G. (1992). On the theoretical status of position class restrictions on inflectional affixes. In *Yearbook of Morphology 1991*, p. 211–241. Springer.

STUMP G. (1998). Inflection. *The handbook of morphology*, p. 13–43.

STUMP G. (2001). *Inflectional morphology : a theory of paradigm structure*. Cambridge University Press.

()