

Décrire la morphologie des verbes en ikota au moyen d'une métagrammaire

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

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

`prenom.nom@univ-orleans.fr`

RÉSUMÉ

Dans cet article, nous montrons comment le concept des métagrammaires introduit initialement par Candito (1996) pour la conception de grammaires d'arbres adjoints décrivant la syntaxe du français et de l'italien, peut être appliquée à la description de la morphologie de l'ikota, une langue bantoue parlé au Gabon. Ici, nous utilisons l'expressivité du formalisme XMG (eXtensible MetaGrammar) pour décrire les variations morphologiques des verbes en ikota. Cette spécification XMG capture les généralisations entre ces variations. Afin de produire un lexique de formes fléchies, il est possible de compiler la spécification XMG, et de sauvegarder le résultat dans un fichier XMG, ce qui permet sa réutilisation dans des applications dédiées.

ABSTRACT

Describing the Morphology of Verbs in Ikota using a Metagrammar

In this paper, we show how the concept of metagrammar originally introduced by Candito (1996) to design large Tree-Adjoining Grammars describing the syntax of French and Italian, can be used to describe the morphology of Ikota, a Bantu language spoken in Gabon. Here, we make use of the expressivity of the XMG (eXtensible MetaGrammar) formalism to describe the morphological variations of verbs in Ikota. This XMG specification captures generalizations over these morphological variations. In order to produce the inflected forms, one can compile the XMG specification, and save the resulting electronic lexicon in an XML file, thus favorising its reuse in dedicated applications.

MOTS-CLÉS : Métagrammaire, morphologie, ikota.

KEYWORDS: Metagrammar, Morphology, Ikota.

1 Introduction

Les langues bantoues (ou bantu) forment une vaste famille de langues africaines. Dans cette famille, le chichewa et le (ki)swahili ont déjà fait l'objet de nombreuses études et sont utilisées comme étalons pour juger de l'expressivité et de la pertinence de théories morphologiques (Mchombo, 1998; Stump, 1992, 1998, 2001) ainsi que de leur implémentation (Roark et Sproat, 2007).

L'ikota (B25) est une langue assez peu décrite du Gabon et de la République Démocratique du Congo. Langue du peuple Kota, avec un nombre de locuteurs estimé à 25000 au Gabon (Idiata, 2007), l'ikota est menacé d'extinction principalement en raison de l'influence du français (langue officielle du Gabon).

Cette langue manifeste de nombreux traits partagés par les langues bantoues (Piron, 1990; Magnana Ekoukou, 2010) :

– l'ikota est une *langue tonale* avec deux tons (Haut et Bas) :

- (1) a. ikàká "famille"
- b. ikákà "paume"
- (2) a. nkúlá "année"
- b. nkúlà "pygmée"

– L'ikota a dix *classes nominales* (les numéros des classes dans le Tableau ci-dessous correspondent à la numérotation de Meinhof) :

TABLE 1 – Classes nominales de l'ikota

classe nominale	préfixe	allomorphes
CL 1	mò-, Ø-	mw-, ñ-
CL 2	bà-	b-
CL 3	mò-, Ø-	mw-, ñ-
CL 4	mè-	
CL 5	ì-, t̥-	dy-
CL 6	mà-	m-
CL 7	è-	
CL 8	bè-	
CL 9	Ø-	
CL 14	ò-, bò-	bw

– l'ikota a un *accord généralisé dans le SN* :

- (3) b-àyitò bá-nèni b-á Ø-mbókà bà-té b-à-t̥á
 Cl.2-femmes Cl.2-grosses Cl.2-du Cl.9-village Cl.2-DEM Cl.2-Présent-mangent

"Ces grosses femmes du village mangent"

Dans cet article, nous ne traitons que la morphologie verbale.

Production d'un lexique de formes fléchies. Notre intention est double : premièrement de fournir une description formelle de la morphologie des verbes en ikota ; deuxièmement, de dériver automatiquement à partir de cette description un lexique de formes fléchies. Dans ce but, nous proposons d'adopter le concept de métagrammaire qui fut introduit par (Candito, 1996) et utilisé pour décrire la syntaxe de langues Indo-Européennes, telles que le français, l'anglais or l'italien. Les grammaires d'arbres lexicalisées à large couverture pour des langues naturelles sont très larges et sont extrêmement gourmandes en ressources pour leur développement et leur

maintenance. Pour cette raison, elles sont souvent produites automatiquement par un logiciel à partir d'une description formelle hautement modulaire appelée une métagrammaire. Celle-ci est beaucoup plus facile à développer et à maintenir.

Nous proposons d'adopter une stratégie similaire pour capturer les généralisations morphologiques des verbes en ikota. Le plan de l'article est le suivant. En Section 2, nous donnons une présentation détaillée de la morphologie des verbes en ikota. En Section 3, nous présentons XMG (*eXtensible MetaGrammar*), un langage formel utilisé pour décrire et combiner des fragments descriptifs réutilisables. En Section 4, nous montrons comment utiliser le cadre offert par XMG pour décrire la morphologie des verbes en ikota. Concrètement, nous présentons une métagrammaire des verbes en ikota que nous avons également codée dans le langage XMG, et qui produit un lexique de formes verbales fléchies. Finalement, en Section 5, nous présentons nos conclusions et perspectives.

2 Les verbes en ikota

Les verbes sont constitués d'une racine lexicale (RV) et de plusieurs affixes de chaque côté de la RV. La description linguistique de l'ikota elle-même étant encore en cours, nous nous focaliserons ici sur les formes verbales de base, en laissant de côté ici les marqueurs de Mode et Voix ainsi que la négation.

Décrivons maintenant les formes infinitives et les trois classes verbales de l'ikota.

Les verbes en ikota sont répartis en trois classes qui dépendent de la forme des marqueurs de l'Aspect et de l'Actif. L'infinitif en ikota constitue une classe hybride. Il est composé d'un préfixe de classe nominale (classe 14) et d'un élément verbal (RV+Prog+Actif).

- (4) a. bòçákà “manger”
 b. bòwéʈfè “donner”
 c. bòbónókò “choisir”

Les exemples (4) illustrent les trois classes verbales.

Il semble toutefois, que le suffixe (Prog+Actif) ait une forme sous-jacente V_{kV}. Dans la variante de l'ikota parlée à Makokou, /k/ est réalisé par [ʈ] quand la voyelle est [ɛ]. Dans la variété dite "standard" de l'ikota, la forme est ékè.

Au niveau sous-jacent, la structure du suffixe infinitival se réduit à AKA, avec trois réalisations de surface distinctes ákà, éʈfè, ókò.

Les exemples ci-dessous illustrent la conjugaison de bòçákà “manger”, un exemple typique de la classe verbale *aka* (dans les exemples ci-dessous, on utilise les 4 valeurs suivantes de *proximité*, *imminent*, *jour*, *proche*, *distant*) :

- (5) m-à-ç-á ñlési
 1sg-Présent-manger-Actif riz
 “Je mange du riz” (Présent)

- (6) a. m-à-ɕ-á-ná yàná
1sg-Passé-manger-Actif-Jour hier
“J’ai mangé hier” (Passé (hier))
- b. m-à-ɕ-á-sá kúlá mwáyèkànàmwé
1sg-Passé-manger-Actif-Distant année dernière
“J’ai mangé l’année dernière” (Passé Distant)
- c. m-é-ɕ-á òlèsì
1sg-Passé-manger-Actif riz
“J’ai mangé du riz” (Passé Récent)
- (7) a. m-é-ɕ-àk-à òlèsì
1sg-Futur-manger-Inaccompli-Actif riz
“Je mangerai du riz” (Futur Moyen)
- b. m-é-ɕ-àk-à-ná yàná
1sg-Futur-manger-Inaccompli-Actif-Jour demain
“Je mangerai demain” (Futur (demain))
- c. m-é-ɕ-àk-à-sá kúlá mwáyàkàmwé
1sg-Futur-manger-Inaccompli-Actif-Distant année prochaine
“Je mangerai l’année prochaine” (Futur Distant)
- d. m-ábí-ɕ-àk-à òsátè
1sg-Futur-manger-Inaccompli-Actif bientôt
“Je mangerai bientôt” (Futur Imminent)

Comme on peut le déduire des exemple ci-dessus, l’ordre des affixes verbaux de l’ikota peut être défini en terme de position. De gauche à droite :

- la classe des préfixes d’accord du Sujet occupe la position la plus à gauche, au début du mot.
- les préfixes de Temps (ou ce que l’on peut plus ou moins identifier comme ayant rapport au Temps) apparaissent à la gauche de la RV.
- le marqueur du progressif est immédiatement à droite de la RV.
- le suffixe de l’Actif occupe le champ à la gauche du Proximal. Il a deux valeurs : Actif et Passif. l’Applicatif et le Causatif seront traités dans des travaux ultérieurs.
- le suffixe du Proximal/Distal occupe la position la plus à droite.

Le Tableau 3 donne le schéma formé par la RV et ses affixes, et le Tableau 2 illustre ce schéma pour bòḥákà “manger”.

3 eXtensible MetaGrammar

eXtensible MetaGrammar (XMG) (MétaGrammaire eXtensible) fait référence à la fois à un langage formel (du type langage de programmation) et à un logiciel, appelé compilateur, qui

TABLE 2 – Formes verbales de bòçákà "manger"

Sujet	Temps	RV	Aspect	Actif	Prox.	Valeur
m-	à-	ç̣		-á		présent
m-	à-	ç̣		-á	-ná	passé, hier
m-	à-	ç̣		-á	-sá	passé distant
m-	é-	ç̣		-á		passé récent
m-	é-	ç̣	-àk	-à		futur moyen
m-	é-	ç̣	-àk	-à	-ná	futur, demain
m-	é-	ç̣	-àk	-à	-sá	futur distant
m-	ábí-	ç̣	-àk	-à		futur imminent

TABLE 3 – Formation du verbe

Sujet-	Temps-	RV	-(Aspect)	-Actif	-(Proximal)
--------	--------	----	-----------	--------	-------------

traite les descriptions écrites dans le langage XMG (Crabbé et Duchier, 2004).

XMG est normalement utilisé pour décrire des grammaires d'arbres lexicalisées. En d'autre mots, une spécification XMG est une description déclarative de structures arborées qui composent la grammaire. Contrairement aux approches antérieures des métagrammaires (notamment (Candito, 1996)), une caractéristique importante du langage XMG est sa déclarativité. XMG offre ainsi au linguiste un langage simple d'utilisation. Concrètement, une description XMG s'appuie sur quatre concepts principaux : (1) **abstraction** : la capacité de donner un nom à un contenu, (2) **contribution** : la capacité à accumuler des informations dans n'importe quel niveau de description linguistique, (3) **conjonction** : la capacité de combiner des éléments d'information, (4) **disjonction** : la capacité de sélectionner de manière non-déterministe des éléments d'information.

Formellement, on peut définir une spécification XMG ainsi :

$$\begin{aligned}
 \text{Règle} & := \text{Nom} \rightarrow \text{Contenu} \\
 \text{Contenu} & := \text{Contribution} \mid \text{Nom} \mid \\
 & \quad \text{Contenu} \vee \text{Contenu} \mid \text{Contenu} \wedge \text{Contenu}
 \end{aligned}$$

Une abstraction est exprimée par une règle de réécriture qui associe un *Contenu* avec un *Nom*. Un tel contenu est soit la *Contribution* d'un fragment de description linguistique (p.e. un fragment d'arbre contribué à la description de la syntaxe), ou une abstraction existante, ou une conjonction ou disjonction de contenus.

Une abstraction en particulier, doit être spécifiquement identifiée comme l'axiome de la métagrammaire. Le compilateur XMG part de cet axiome et utilise les règles de réécriture pour produire une dérivation complète. Quand une disjonction est rencontrée, elle est interprétée comme offrant plusieurs alternatives pour continuer : le compilateur explore successivement chaque alternative. De cette manière, l'exécution d'une métagrammaire produit typiquement de nombreuses dérivations. Le long d'une dérivation, les contributions sont simplement accumulées

de manière conjonctive. À la fin de la dérivation, cette accumulation de contributions est interprétée comme une spécification et donnée à un résolveur pour produire des structures solutions. La collection de toutes les structures produites de cette manière forme la grammaire résultante. Celle-ci peut être inspectée grâce à un outil graphique, ou bien exportée au format XML.

Le compilateur XMG est disponible librement sous une licence compatible avec la GPL, et est fourni avec une documentation raisonnable.¹ Il a été utilisé pour concevoir, entre autres, de vastes grammaires arborées pour le français (Crabbé, 2005; Gardent, 2008), l'anglais (Alahverdzhieva, 2008) et l'allemand (Kallmeyer *et al.*, 2008).

XMG a été spécifiquement conçu pour écrire des grammaires arborées, hautement modulaires, à large couverture, couvrant à la fois l'expression syntaxique et le contenu sémantique. Bien qu'XMG n'ait jamais été prévu pour exprimer la morphologie, notre projet (travail en cours) démontre qu'il peut-être réutilisé facilement pour cette tâche, tout du moins dans le cas d'une langue agglutinante comme l'ikota.

4 Métagrammaire de la morphologie verbale de l'ikota

Notre formalisation de la morphologie verbale de l'ikota s'inspire du modèle Paradigm-Function Morphology (Stump, 2001) qui repose sur le concept de classes de positions. Plus précisément, nous empruntons la notion de *domaine topologique* à la tradition de la syntaxe descriptive de l'allemand (Bech, 1955) pour instancier ces classes. Un domaine topologique consiste en une séquence linéaire de champs. Chaque champ peut accueillir des contributions, et il peut y avoir des restrictions sur le nombre d'items qu'un champ peut ou doit recevoir. Dans notre cas, le domaine topologique d'un verbe sera tel que décrit dans le Tableau 3, et chaque champ accueillera au plus un item, où chaque item est la *forme phonologique lexicale*² d'un morphème.

Blocs élémentaires. La métagrammaire est exprimée au moyen de blocs élémentaires. Un bloc contribue simultanément à 2 dimensions de descriptions linguistiques : (1) la phonologie lexicale : contributions aux champs du domaine topologique, (2) la flexion : contribution de traits morphosyntaxiques. Par exemple :

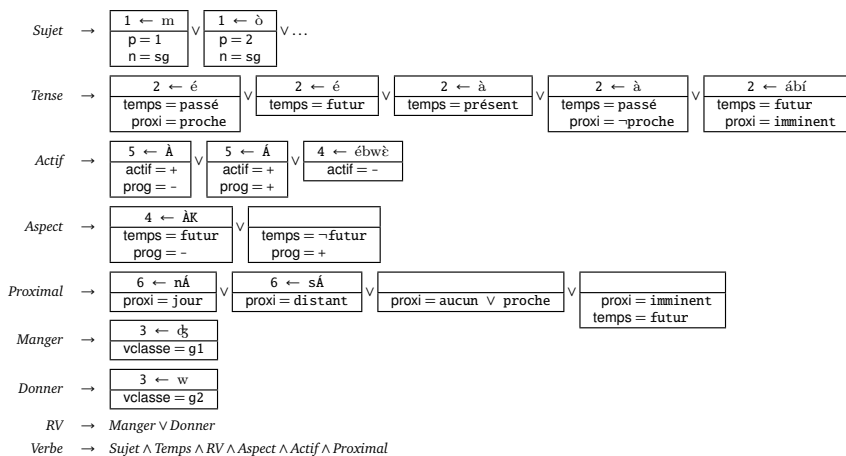
2 ← é
temps = passé
proxi = proche

contribue é au champ numéro 2 du domaine topologique, et les traits temps = passé et proxi = proche à la flexion. Les contributions de traits provenant de différents blocs sont unifiées : de cette manière, la dimension de flexion sert également comme un niveau de médiation et de coordination durant l'exécution de la métagrammaire. Comme le Tableau 2 l'illustre clairement, la morphologie de l'ikota n'est pas proprement compositionnelle : en effet, les contributions sémantiques des morphèmes sont déterminées au travers d'une coordination de contraintes mutuelles dans le niveau de flexion.

1. Voir <http://sourcesup.cru.fr/xmg>

2. Nous adoptons ici la perspective à 2 niveaux qui distingue phonologie lexicale et de surface (Koskenniemi, 1983)

FIGURE 1 – Métagrammaire de la morphologie verbale de l'ikota



Les traits morphosyntaxiques. Nous utilisons *p* et *n* pour *personne* et *nombre* ; *temps* avec pour valeurs possibles *passé*, *présent*, et *futur* ; *proxi* pour le *marqueur proximal* (*aucun*, *imminent*, *jour*, *proche*, *distant*) ; *vclasse* pour la classe verbale (*g1*, *g2*, *g3*) ; et deux traits polaires : *actif* pour la *voix* et *prog* pour l'*aspect progressif* : *prog=-* marque un évènement en déroulement.

Signes phonétiques lexicaux. Une étude attentive des données disponibles sur l'ikota suggère que l'on peut mieux rendre compte des régularités parmi les classes verbales en introduisant une voyelle *lexicale A* qui est réalisée, au niveau surfacique, par *a* pour *vclasse=g1*, *ɛ* pour *vclasse=g2*, et *ɔ* for *vclasse=g3*, et une consonne *lexicale K* qui est réalisée par *tʃ* pour *vclasse=g2*, et *k* sinon.

Règles. La Figure 1 montre un fragment de notre métagrammaire préliminaire de la morphologie verbale de l'ikota. Chaque règle définit comment une abstraction peut être réécrite. Par exemple *Temps* peut être réécrit par un bloc quelconque représentant une disjonction de 5 blocs. Pour produire le lexique des formes fléchies décrites par notre métagrammaire, le compilateur XMG calcule toutes les réécritures non-déterministes possibles en partant de l'abstraction *Verbe*.

Exemple de dérivation. Considérons comment *óçàkàná* (*(demain), tu mangeras*) est dérivé par notre système formel en partant de l'abstraction *Verbe*. Premièrement, *Verbe* est remplacé par *Subjet ∧ TempsRV ∧ Aspect ∧ Actif ∧ Proximal*. Puis chaque élément de cette conjonction logique

FIGURE 2 – Une dérivation avec succès

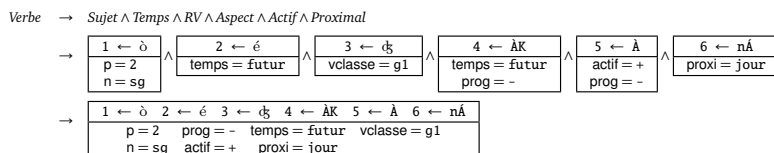
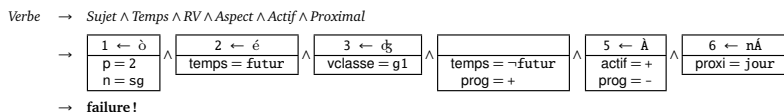


FIGURE 3 – Une dérivation avec échec : conflits sur temps et prog



(l'ordre est sans importance) est, à son tour, remplacé. Par exemple, *Sujet* est alors remplacé par un bloc de la disjonction correspondante : le compilateur XMG essaie toutes les possibilités ; l'une d'entre elles choisira le 2ème bloc. La Figure 2 montre l'étape initiale, une étape au milieu, et l'étape finale de la dérivation. La phonologie lexicale de l'entrée lexicale résultante est obtenue en concaténant, dans l'ordre linéaire du domaine topologique, les items contribués aux différents champs ici : ò+é+çş+ÀK+À+nÁ.

La Figure 3 montre un exemple d'une dérivation rencontrant un échec, donc, qui ne mène pas à la production d'une entrée du lexique. L'échec est dû à des valeurs contradictoires pour les traits temps (futur et ¬futur) et aussi prog (+ et -).

Phonologie de surface. Pour l'instant, notre métagrammaire modélise uniquement le niveau lexical de la phonologie. Le niveau surfacique peut en être dérivé par post-traitement. Pour notre exemple, puisque $vclasse=g1$, le A lexical devient a en surface, et le K devient k. Ainsi nous obtenons : ò+é+çş+àk+à+ná, et finalement (par effacement de voyelle) óçşàkàná.

L'approche de XMG basée sur les contraintes en fait une plateforme idéale pour l'intégration, par exemple, de la *phonologie à deux niveaux* puisque celle-ci est précisément une contrainte entre la phonologie lexicale et surfacique (Koskeniemi, 1983). Cette extension de XMG fait partie de la feuille de route d'une thèse en cours.

Réserves. Notre formalisation de la morphologie de l'ikota est encore au stade préliminaire. Au fur et à mesure que nous progressons, des questions apparaissent pour lesquelles nous n'avons pas encore suffisamment de données. Par exemple, il est aisé de déduire de la Figure 1 que notre métagrammaire (délibérément) omet le "futur passif" ; de nouvelles données venant de locuteurs

natifs permettront de valider ou non son existence.

Il est également trop tôt pour nous pour, ne serait-ce qu'esquisser une formalisation du système tonal de l'ikota, et ses implications sur les contours prosodiques des formes verbales. Par conséquent, et dans l'intérêt d'une morphologie descriptive exacte, nous avons été amenés à adopter certaines astuces, dans notre description formelle, comme un recours pratique plutôt que par positionnement théorique : c'est ainsi le cas de l'alternance tonale à la voix active.

5 Conclusions et perspectives

Dans cet article, nous avons proposé une description formelle, quoique préliminaire, de la morphologie verbale de l'ikota, une langue africaine peu dotée et dont la description fine n'est pas achevée. Cette description utilise un langage de haut niveau permettant une réflexion linguistique sur la redondance de la représentation morphologique. Ce faisant, nous avons illustré comment l'approche métagrammaticale peut contribuer de manière utile au développement de ressources lexicales électroniques.

Ainsi, à partir de cette description, à l'aide du compilateur XMG, nous produisons automatiquement un lexique de formes verbales fléchies avec leurs traits morphosyntaxiques. Ce lexique peut être exporté au format XML, offrant une ressource normative facilement réutilisable pour cette langue sous-dotée.

D'un point de vue méthodologique, l'utilisation de XMG nous a permis de tester rapidement nos intuitions linguistiques en générant toutes les formes verbales prédites et leurs traits, et donc de valider ces résultats au regard des données disponibles.

Un autre avantage d'adopter l'approche par métagrammaire est que, grâce au même outil (formel et logiciel), nous serons en mesure de décrire également la syntaxe de cette langue à l'aide d'une grammaire d'arbres adjoints, ce qui constitue le sujet d'une thèse en cours.

Références

- ALAHVERDZHIEVA, K. (2008). XTAG using XMG. Master Thesis, Nancy Université.
- BECH, G. (1955). *Studien über das deutsche Verbum infinitum*. Det Kongelige Danske videnskabelnes selskab. Historisk-Filosofiske Meddelelser, bd. 35, nr.2 (1955) and bd. 36, nr.6 (1957). Munksgaard, Copenhagen. 2nd unrevised edition published 1983 by Max Niemeyer Verlag, Tübingen (Linguistische Arbeiten 139).
- CANDITO, M. (1996). A Principle-Based Hierarchical Representation of LTAGs. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING'96)*, volume 1, pages 194–199, Copenhagen, Denmark.
- CRABBÉ, B. (2005). *Représentation informatique de grammaires fortement lexicalisées : Application à la grammaire d'arbres adjoints*. Thèse de doctorat, Université Nancy 2.
- CRABBÉ, B. et DUCHIER, D. (2004). Metagrammar redux. In CHRISTIANSEN, H., SKADHAUGE, P. R. et VILLADSEN, J., éditeurs : *Constraint Solving and Language Processing, First International Workshop (CSLP 2004), Revised Selected and Invited Papers*, volume 3438 de *Lecture Notes in Computer Science*, pages 32–47, Roskilde, Denmark. Springer.

- GARDENT, C. (2008). Integrating a Unification-Based Semantics in a Large Scale Lexicalised Tree Adjoining Grammar for French. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 249–256, Manchester, UK. Coling 2008 Organizing Committee.
- IDIATA, D. F. (2007). *Les langues du Gabon : données en vue de l'élaboration d'un atlas linguistique*. L'Harmattan.
- KALLMEYER, L., LICHTÉ, T., MAIER, W., PARMENTIER, Y. et DELLERT, J. (2008). Developing a TT-MCTAG for German with an RCG-based Parser. In *The sixth international conference on Language Resources and Evaluation (LREC 08)*, pages 782–789, Marrakech, Morocco.
- KOSKENNIEMI, K. (1983). *Two-Level Morphology : a general computational model for word-form recognition and production*. Thèse de doctorat, University of Helsinki.
- MAGNANA EKOUKOU, B. (2010). Morphologie nominale de l'ikota (B25) : inventaire des classes nominales. Mémoire de Master 2, Université d'Orléans.
- MCHOMBO, S. A. (1998). Chichewa : A Morphological Sketch. In SPENCER, A. et ZWICKY, A., éditeurs : *The Handbook of Morphology*, pages 500–520. Blackwell, Oxford, UK & Cambridge, MA.
- PIRON, P. (1990). Éléments de description du kota, langue bantoue du gabon. mémoire de licence spéciale africaine, Université Libre de Bruxelles.
- ROARK, B. et SPROAT, R. (2007). *Computational approaches to morphology and syntax*. Oxford University Press, USA.
- STUMP, G. T. (1992). On the theoretical status of position class restrictions on inflectional affixes. In BOOIJ, G. et van MARLE, J., éditeurs : *Yearbook of Morphology 1991*, pages 211–241. Kluwer.
- STUMP, G. T. (1998). Inflection. In SPENCER, A. et ZWICKY, A. M., éditeurs : *The Handbook of Morphology*, pages 13–43. Blackwell, Oxford & Malden, MA.
- STUMP, G. T. (2001). *Inflectional Morphology : a Theory of Paradigm Structure*, volume 93. Cambridge University Press.