

**EXTRACTION SEMI-AUTOMATIQUE DES CONTEXTES SYNTAXIQUES  
DES EMPRUNTS DU FRANÇAIS AUX LANGUES BURKINABÈ  
DANS ROUGBÉINGA DE NORBERT ZONGO**

**Célestin ZOUMBARA**

Université Norbert Zongo, Burkina Faso

celestin.zoumbara@unz.bf

[celestin.zoumbara@unz.bf](mailto:celestin.zoumbara@unz.bf)

**Résumé :** Dans la perspective distributionnaliste de la langue, le signe linguistique se définit en fonction de ses différents contextes d'apparition dans les énoncés, c'est-à-dire l'environnement linguistique dans lequel le signe apparaît dans l'usage de la langue. Ce faisant, cet article se concentre sur les contextes d'insertion des emprunts dans les textes littéraires burkinabè. Il est construit autour d'une information centrale en lexicologie et en sémantique, l'information définitoire et dénominative, dont il pose des préalables théoriques. La méthode Extranéo (Extraction automatique des néologismes) a été adaptée pour réaliser l'étude. Cette méthode est une combinaison de logiciels de Traitement automatique des langues naturelles (TALN) et de règles en langage de programmation Perl permettant d'extraire les néologismes ainsi que leurs contextes d'insertion dans les énoncés.

**Mots-clés :** contextes syntaxiques, emprunts, Extranéo, français du Burkina Faso

**SEMI-AUTOMATIC EXTRACTION OF SYNTACTIC CONTEXTS OF FRENCH  
BORROWINGS FROM BURKINABÈ LANGUAGES IN ROUGBÉINGA BY  
NORBERT ZONGO**

**Abstract:** In the distributionalist perspective of language, the linguistic sign is defined in terms of its different contexts of appearance in statements. In doing so, this article focuses on the contexts of insertion of borrowings in burkinabè novelistic texts. It is built around a central information in lexicology and semantics, the defining and denominative information, for which it lays the theoretical foundations. In practice, the Extranéo method (Automatic Extraction of Neologisms) was adapted to make the experiences. This method is a combination of automatic Natural Language Processing (NLP) tools and rules in the Perl programming language that allows the extraction of neologisms and their contexts of insertion in the statements.

**Keywords:** syntactic contexts, borrowing, Extranéo, French language of Burkina Faso

## **Introduction**

Le multilinguisme qui caractérise le Burkina est à l'origine de différents types de particularités lexicales ou diatopismes dans le français du Burkina Faso. Les néologismes ou nouveaux mots constituent une caractéristique majeure de ces diatopismes. Ils sont généralement créés suivant deux types de matrices de création lexicale (Sablayrolles, 2017). Il s'agit de matrices internes qui permettent d'utiliser diverses ressources du français pour former des néologismes morpho-sémantiques, syntactico-sémantiques, morphologiques et

phraséologiques, et d'une matrice externe qui repose exclusivement sur l'emprunt. Dans le contexte du français du Burkina Faso, l'emprunt constitue l'un des phénomènes langagiers qui entretient la présence de diatopismes. Ces emprunts proviennent notamment de langues africaines locales, de langues africaines non locales et de langues non africaines (Prignitz, 1994). Ils sont présents aussi bien dans les discours oraux, dans la presse orale ou écrite que dans la littérature. Même si ces unités linguistiques ne sont pas toujours destinées à intégrer le lexique usuel du français parlé au Burkina, elles renferment d'importantes informations socio-culturelles et lexicologiques dont la prise en compte est un gage pour une communication réussie.

À la suite d'études précédentes, qui ont mis en exergue la description de ces unités lexicales dans divers domaines (Ouédraogo, 2000 ; Koama, 2004 ; Daïla, 2017), et dans une perspective lexicologique, nous mettons l'accent sur les contextes d'insertion des emprunts dans les énoncés littéraires. La question qui fonde cette étude est formulée comme suit : Dans quels contextes syntaxiques les lexies empruntées aux langues burkinabè sont-elles insérées dans *Roubêinga* de N. Zongo ? Cette question a suscité une hypothèse qui stipule que dans *Roubêinga* les emprunts aux langues burkinabè sont insérés dans des contextes spécifiques pouvant être extraits en utilisant une méthode semi-automatique de fouille de textes. Dès lors, cette étude vise à mettre en relief les contextes d'insertion des emprunts du français aux langues burkinabè dans *Roubêinga* à partir de l'application d'une méthode semi-automatique de fouille de textes. L'intérêt de cette étude est qu'elle apporte des informations supplémentaires dans la description du français d'Afrique à travers la cohabitation entre le français et les langues burkinabè. Ainsi, il ne s'agit pas d'inventorier les emprunts figurant dans l'œuvre, mais de mettre au point une méthode semi-automatique permettant d'extraire les contextes syntaxiques des emprunts dans le contexte du français du Burkina Faso. Quatre principales sections constituent l'ossature de cet article. La première section est consacrée à une clarification conceptuelle de différents phénomènes langagiers afin de mieux positionner la notion d'emprunt.

Quant à la deuxième section, elle présente la méthode d'analyse appliquée au corpus. La troisième section traite de l'analyse du corpus. Les résultats de l'étude et la discussion sont enfin présentés dans la quatrième section.

## 1. Clarification conceptuelle

S'intéressant au français parlé au Burkina Faso, pays multilingue car comportant cinquante-neuf langues nationales (Kedrebéogo et al., 1988), Kéita et Pardevan (2020 :93) montrent que « plusieurs langues y sont en contact et engendrent des phénomènes langagiers comme le xénisme, les alternances codiques, les interférences linguistiques, les calques et les emprunts ». Cette remarque présente diverses notions linguistiques dont la clarification est nécessaire pour une bonne conduite de la présente étude.

### 1.1. Xénisme, alternance codique

D'après Dubois et al. (2001 :52), le xénisme est « une unité lexicale constituée par un mot d'une langue étrangère et désignant une réalité propre à la culture des locuteurs de cette langue ». S'appuyant sur cette définition, Kéita et Pardevan (2020 :7) montrent que le

xénisme est un phénomène langagier qui « fait la description d'une réalité spécifique qui n'existait pas dans la langue cible et est ainsi le premier stade de l'emprunt ». Ainsi, le xénisme permet de combler un vide lexical et son intégration dans la mémoire collective lui permet d'intégrer le lexique de la langue. De façon précise, le xénisme traduit une réalité culturelle spécifique conduisant le locuteur à l'utiliser comme tel afin de mieux faire passer un message.

Le phénomène de l'alternance codique ou *code switching* (en anglais), quant à lui, consiste, dans un texte rédigé dans une langue L1 ou langue A, à changer de code linguistique pour émettre une séquence de discours. Cette séquence peut être brève ou courte, se limiter à une seule lexie ou s'étendre sur un paragraphe, voire sur un texte entier dans une langue L2 ou langue B. Ce phénomène langagier, d'après Cartier (2019 :1), ne fait pas partie de l'emprunt linguistique « puisqu'il s'agit d'un phénomène dans lequel l'énonciateur assume le passage d'une langue à l'autre, et ne cherche aucunement à intégrer le passage en langue L1 dans la langue L2 ». Le phénomène de l'alternance codique laisse intacte les langues L1 et L2.

### 1.2. *Interférence linguistique, calque*

Le phénomène langagier d'interférence linguistique ou *code mixing* (en anglais) revient à faire un mélange de deux ou plusieurs langues. Ainsi, il intervient lorsqu'un locuteur insère des lexies d'une langue L1 dans une langue L2 afin de véhiculer un message. Cela peut résulter principalement d'un contexte sociolinguistique présentant une cohabitation entre plusieurs langues. C'est dans ce sens que Cartier (2019 :1) souligne que l'interférence linguistique est un phénomène qui « est lié à des situations sociolinguistiques spécifiques de bi- ou multilinguisme, il peut être limité à des groupes sociolinguistiques spécifiques et aboutit parfois à des créoles ». Au nombre des phénomènes langagiers à distinguer de l'emprunt, on peut citer également le calque. Selon Rondeau (1986 :48), le calque est « la traduction littérale et mot à mot d'un terme d'une langue dans l'autre, y compris le modèle syntaxique ». L'élément linguistique considéré comme étant un calque provient d'une langue L1 mais est exactement utilisé en adoptant des éléments caractéristiques de la langue L2. Ce faisant, précisent Kéita et Pardevan :

Le calque est un type d'emprunt particulier dont la forme est authentiquement de la langue d'arrivée, mais le sens de la langue de départ. Il est une traduction littérale de la langue source qui intègre le sens étranger sous une forme nouvelle obtenue par la traduction dans la langue d'accueil.

Kéita et Pardevan (2020 :98)

### 1.3. *Emprunt*

La notion linguistique d'emprunt décrit le résultat de l'insertion d'une unité lexicale d'une langue A dans une langue B. En fonction du mode d'insertion et des caractéristiques des langues en usage, diverses acceptions ont été formulées pour décrire la notion d'emprunt. Aussi convient-il pour nous de préciser l'acception que nous prenons en compte dans cette étude afin de lever toute ambiguïté. Selon Dubois et al. (2001 :177), il y a emprunt linguistique quand un parler A utilise et finit par intégrer une unité ou un trait linguistique

qui existait précédemment dans un parler B (dit langue source) et que A ne connaissait pas ; l'unité ou le trait emprunté sont eux-mêmes qualifiés d'emprunts. On peut relever deux états descriptifs de l'emprunt linguistique : l'utilisation d'une lexie et, en fin de compte, son intégration dans le lexique de la langue A. Chez Loubier (2011 :10), c'est essentiellement le deuxième état qui est caractéristique de l'emprunt, car il définit l'emprunt comme étant un procédé par lequel « les utilisateurs d'une langue adoptent intégralement, ou partiellement, une unité ou un trait linguistique (lexical, sémantique, phonologique, syntaxique) d'une autre langue ». Pour lui, la lexie considérée comme étant un emprunt doit avoir été adoptée par la langue A. Toutefois, cet état suppose le premier état évoqué par Dubois et al. En effet, d'après Cartier (2019 :9), trois cycles de vie sont caractéristiques des emprunts, c'est-à-dire « l'émergence, la diffusion et l'institutionnalisation ». En tenant compte de ces trois phases de vie des emprunts, nous convenons avec Grevisse et Goosse (2011 :160) que les emprunts sont « les éléments qu'une langue au cours de son histoire a pris à d'autres langues ». Cette catégorie de lexies concerne donc les termes provenant d'une langue étrangère, vivante ou ancienne. Notre étude s'inscrivant dans le contexte du français parlé au Burkina Faso, nous considérons comme emprunt les lexies empruntées à d'autres langues et insérées dans *Rougbeïnga* de l'écrivain burkinabè N. Zongo. Cette considération rejoint la déduction de Natama (2021) selon laquelle l'auteur de l'œuvre étant un utilisateur du français, les lexies issues d'autres langues et utilisées dans le roman sont considérées comme des emprunts. Au regard du contexte sociolinguistique du Burkina Faso, caractérisé par un multilinguisme ethnique, les emprunts du français, d'après Prignitz (1994) proviennent de trois principales sources : Les langues africaines locales, c'est-à-dire les langues burkinabè (bwamou, dioula, moré, san, goulmacéma), les langues africaines non locales (ashanti, wolof, yorouba) et les langues non africaines (anglais, allemand, chinois).

#### 1.4. Contexte syntaxique

La notion de contexte, selon Gross (2010 :187), constitue l'une des plus importantes de l'analyse linguistique « dans la mesure où presque tous les prédicats sont polysémiques et que seule la distribution permet de déterminer la lecture adéquate ». Cette notion tire sa source du structuralisme, né à partir des travaux du linguiste américain Bloomfield avec la publication de son ouvrage *Language* (1935). Dans cet ouvrage, il insiste notamment sur la segmentation de l'énoncé linguistique en unités plus petites. La phrase est ainsi segmentée en ses constituants immédiats jusqu'aux morphèmes. Ce qui permet d'en étudier la distribution. C'est cette conception qui donnera naissance au distributionnalisme. Ainsi, la compréhension et la description des lexies reposent en partie sur la somme de leurs contextes d'utilisation. Considérons, par exemple, l'extrait suivant : « La puissance de sa magie, son invulnérabilité et surtout son don d'ubiquité avaient fait de lui un être exceptionnel que les Mossi désignaient sous l'appellation de *homme-force* » (Zongo, 2009, p. 96). Dans cet extrait, le néologisme par composition « homme-force » est inséré pour réaliser une dénomination. L'insertion est réalisée à l'aide de la glose *désignaient sous l'appellation de*.

Dans un texte, le contexte implique que l'unité de base ne soit pas la lexie, mais la phrase. Ce qui revient à prendre en compte l'environnement dans lequel la lexie est employée à partir d'une analyse basée sur la disposition syntaxique des composantes de la

phrase. Il ne s'agit pas simplement d'identifier le passage textuel dans lequel le néologisme est inséré. C'est ce que Gross (2010) relate en montrant que le contexte ne constitue pas une délimitation mécanique que l'on peut effectuer en sélectionnant un nombre déterminé de mots à gauche et à droite mais relève d'une analyse qui met en jeu la totalité des informations syntaxiques qui environnent la lexie. Ce faisant, et en partant de la perspective distributionnelle de Harris (1968), nous retenons, dans le cadre de cette étude, que le contexte syntaxique d'une lexie désigne l'ensemble des lexies qui entretiennent avec elle une relation de dépendance syntaxique. Ce contexte peut être identifiable à l'aide de méthodes de fouilles de textes basées sur la glose. D'après Mela et al. (2011 :9), « les gloses sont des commentaires en situation parenthétique, souvent introduits par des marqueurs ». Ces marqueurs peuvent être lexicaux comme *appelé, signifiant, désignant, c'est-à-dire, ou, comme* ou typographiques, à travers l'utilisation d'une virgule, des parenthèses, des deux points, etc. Les marqueurs de glose signent une relation de sémantique lexicale et peuvent introduire, dans un énoncé, un contexte définitoire ou un contexte de dénomination. Le contexte définitoire, selon Cartier (2011 :619), « rassemble les informations les plus essentielles du sens des mots et des catégories d'objets qu'ils désignent ».

Quant au contexte de dénomination, précise Mortureux (1984), il renvoie à la nécessité de donner un nom à un objet, à un lieu, à un concept nouveau. Dans le cadre de cette étude, nous avons choisi d'adapter la méthode Extranéo pour réaliser l'extraction des contextes syntaxiques des emprunts.

## 2. Approche d'extraction automatique des emprunts : la méthode Extranéo

Nous avons montré que bon nombre de raisons favorisent l'utilisation d'emprunts dans la pratique quotidienne d'une langue donnée. Les emprunts constituent ainsi un élément majeur dans la description d'un état synchronique de langue et permettent non seulement d'enrichir cette langue, mais aussi d'attester de sa vitalité. Dans ce sens, l'extraction des emprunts est une entreprise permettant d'appréhender les lexies qu'une langue à emprunter à d'autres langues en mettant en relief, entre autres caractéristiques, les domaines socio-culturels concernés par les emprunts ainsi que les catégories grammaticales auxquelles ils renvoient. Tout comme pour les néologismes d'une façon globale, l'extraction des emprunts repose sur deux types d'approches. Une approche manuelle et une approche automatique. L'approche manuelle s'appuie sur un protocole visant à faire un recensement manuel des lexies obtenues par emprunt dans une production linguistique donnée. Pour ce faire, l'on s'appuie sur un (des) dictionnaire(s) pris comme corpus d'exclusion, c'est-à-dire une référence qui permet, comme le soulignent Cartier et Sablayrolles (2009 :172), de décider d'un « certificat de néologisme » pour une unité lexicale soupçonnée de nouveauté ». Cette approche nécessite un long temps de travail avec des limites conséquentes, car étant soumise à des conditions physiologiques comme la fatigue et l'oubli.

Quant à l'approche automatique, elle consiste à configurer les règles d'extraction afin de mettre en place un système automatique exécuté via un ordinateur. Ce faisant, elle peut être appliquée à un gros corpus et ramène les requêtes exécutées en un laps de temps. Dans cette perspective, visant l'exploitation de corpus de plusieurs dizaines de millions de mots dans la description du français du Burkina Faso, nous avons mis au point la méthode

Extranéo (Extraction automatique des néologismes), qui comporte quatre principales étapes (Zoumbara et al., 2020). Ainsi, une contribution de cette étude est de montrer que la méthode Extranéo peut être adaptée spécifiquement pour extraire les emprunts du français et leurs contextes syntaxiques à partir de données textuelles en français du Burkina, en partant du fait que les catégories grammaticales caractéristiques des emprunts sont les adjectifs, les noms et les verbes (Kéita et Pardevan, 2020).

### 3. Analyse du corpus de l'étude

Dans cette section, nous présentons le corpus de l'étude ainsi que la méthode d'analyse mise en œuvre. Ainsi, nous déclinons quelques caractéristiques du corpus, dans un premier temps, et, dans un second temps, nous décrivons les mesures d'évaluation qui nous ont permis de quantifier la représentativité des contextes syntaxiques étudiés.

#### 3.1. Corpus

En termes de contenu, le corpus étudié est un roman qui porte sur la période coloniale et raconte l'histoire de Soura Dakuo (un Bwaba), alias Roughbèinga et de Balily Bapio (un Gourounsi). Les deux s'évadent d'un camp de travaux forcés depuis Bamako au Mali et retournent sur leur terre natale afin d'organiser une lutte armée contre le colon français. L'histoire se déroule à un moment où, un effort surhumain est demandé aux colonies pour soutenir la métropole engagée dans la seconde guerre mondiale. Plusieurs exactions du colon sont relatées dans le roman, allant du pillage des matières premières au massacre des populations locales en passant par le vol, le viol, les tortures, etc.

Dans la narration du récit, Norbert ZONGO emprunte des lexies à des langues burkinabè, en l'occurrence le bwamou et le moré. Dans cet article, nous nous intéressons principalement aux emprunts appartenant à la catégorie grammaticale des noms. D'après Natama (2021), ils sont insérés pour désigner des réalités comme les êtres vivants, les nourritures locales, les objets fabriqués, les éléments de la nature, les pratiques culturelles, etc. À titre illustratif, les deux extraits ci-dessous montrent un emprunt au bwamou, *timbowni* « grand tam-tam » désignant un objet fabriqué se rapportant à des pratiques culturelles, et deux emprunts au moré *Zouri* « chenille » et *Zouanga* « mouche », désignant des êtres vivants :

- Lorsque pour l'imminence d'un grand danger, le **timbowni** (grand tam-tam) grondait, et que sa voix à grandes enjambées escaladait les collines, tous les habitants des kuyis environnants regagnaient Houndé leur capitale, qui pouvait tenir un siège de plusieurs lunes. (page 8)
- Lorsqu'au sein d'un peuple, certains éléments tentent de vivre comme **Zouri** la chenille ou **Zouanga** la mouche, ce peuple est condamné à mort tôt ou tard. (page 14)

À partir du format électronique de l'œuvre, nous avons constitué un sous-corpus regroupant les énoncés dans lesquels les emprunts ont été insérés. La constitution de ce sous-corpus s'est passée en 2 phases. La première phase a consisté en une lecture du roman afin de sélectionner les passages dans lesquels des emprunts au bwamou et au moré ont été insérés. La deuxième phase a consisté à rassembler les passages sélectionnés, allant d'un

simple énoncé au paragraphe, dans un fichier Word (.docx) converti ensuite en fichier texte brut (.txt) pour constituer un sous-corpus de 1 281 mots. C'est ce sous-corpus qui a été soumis à la méthode Extranéo ; l'étiquetage a été réalisé à partir de l'étiqueteur morphosyntaxique TreeTagger (Schmid, 1995), dont le jeu d'étiquettes est présenté en annexe. De façon spécifique, après l'étape de sélection manuelle des néologismes validés, nous avons séparé les néologismes issus de la matrice interne de ceux issus de la matrice externe, c'est-à-dire les emprunts. Ceux-ci sont ensuite utilisés pour concevoir des patrons syntaxiques. Le terme « patron », dérivé de l'anglais « pattern » se traduisant par « modèle ou structure », consiste à définir le modèle d'une structure linguistique donnée par « des indices de surface » ; cela nécessite la connaissance à priori des informations recherchées (Poibeau & Nazarenko, 1999). Dans notre cas précis, les indices de surface sont des marqueurs de glose lexicaux ou typographiques. Ainsi, nous appréhendons le patron syntaxique comme un modèle de structure syntaxique qui se veut répétitive dans un corpus textuel donné et constituant une source d'informations diverses. Les patrons conçus sont ensuite exécutés sur le corpus à travers le logiciel de textométrie TXM (Heiden et al., 2010). Ce logiciel intègre un concordancier permettant d'afficher le contexte d'utilisation de la forme spécifiée. Un concordancier est un programme, qui, sur la base d'une lexie précise appelée pivot, permet de délimiter le contexte ou l'environnement syntaxique donné pour chaque occurrence du pivot tout en organisant les extraits par un tri afin de faciliter le dépouillement (Pincemin et al., 2006). Pour exécuter des requêtes via l'interface de ce logiciel, le langage CQP (Evert & Hardie, 2011) a été utilisé afin de concevoir les patrons morphosyntaxiques. Ce langage consiste à utiliser des expressions régulières pour formuler des requêtes qui sont exécutées par le moteur de recherche sur le corpus. Les requêtes formulées sont toujours placées entre crochets. La principale caractéristique de ce langage, d'après Evert & Hardie (2011), est qu'il permet d'utiliser les expressions régulières à deux niveaux. Premièrement, la requête formulée peut porter sur une chaîne de caractères précise, c'est-à-dire une forme lexicale donnée, ou sur une valeur d'annotation définie par l'utilisateur. Ainsi, la requête visant à rechercher les occurrences de la lexie « papillon » dans un corpus sera formulée ainsi : [word ="papillon"]. Deuxièmement, la requête peut porter sur une séquence (début ou fin) d'une lexie. Par exemple, pour rechercher toutes les lexies terminées par « -illon », la requête sera : [word =".\*illon"]. Cette spécification peut également s'appliquer à la recherche des lexies par catégorie grammaticale. Ainsi, en exécutant par exemple la requête [frpos="NOM"], l'utilisateur demande au logiciel d'afficher toutes les unités lexicales dont l'étiquette attribuée par la version française de TreeTagger (frpos) est « nom commun » (NOM).

### **3.2. Mesures d'évaluation**

Les résultats obtenus après l'exécution des patrons syntaxiques ont ensuite fait l'objet d'une évaluation en fonction de chaque contexte syntaxique afin d'en évaluer la représentativité. Pour cela, nous nous sommes appuyé sur deux notions utilisées généralement en système de recherche d'informations (Pinel-Sauvagnat & Mothe (2013). Il s'agit des notions de bruit et de silence documentaires. Dans une recherche d'information

donnée, la notion de bruit documentaire fait référence aux documents restitués à l'utilisateur mais ne répondant pas à son besoin.

Quant au silence documentaire, il fait référence aux documents pertinents existants mais que le système n'a pas restitués à l'utilisateur. La prise en compte de ces notions permet de calculer trois mesures, encore appelées scores, afin de quantifier la qualité du système utilisé dans le processus de recherche d'informations. Il s'agit des scores de précision et de rappel (Kent et al., 1955), ainsi que du score de F-mesure (Rijsbergen, 1979), qui est une moyenne harmonique entre la précision et le rappel. Ces différents scores obtiennent généralement une valeur variant entre 0 et 1. La valeur de 1 désigne une valeur optimale. Le score de précision quantifie la proportion de documents pertinents (Pert) dans l'ensemble de documents restitués (Rest) conformément à l'Équation 1. Dans notre cas précis, il s'agit de calculer le rapport entre le nombre de contextes intégrant un emprunt et le nombre de contextes retournés par les patrons syntaxiques. Quant au rappel, il quantifie la proportion de documents pertinents restitués suivant l'opération posée à travers l'Équation 2, c'est-à-dire que ce score calcule le rapport entre le nombre de contextes intégrant un emprunt et le nombre d'emprunts réel. Si, pour une requête donnée, la valeur de la précision vaut 1, cela signifie que le système utilisé identifie correctement tous les documents pertinents dans l'ensemble de documents restitués. Si la valeur de rappel vaut 1, cela signifie que tous les résultats restitués sont pertinents. À titre illustratif, une requête du patron syntaxique « NOM + DET : ART + NOM » dans un corpus qui ramène uniquement les occurrences de ce patron passe sous silence des structures comme « NOM + ADJ + DET : ART + NOM » ou encore « NOM + PREP + DET : ART + NOM ». Un tel résultat produit un silence et fait baisser la valeur du rappel. Par contre, si pour pallier cette situation l'utilisateur recherche uniquement la catégorie grammaticale « NOM », le système ramènera tous les noms contenus dans le corpus même s'ils ne s'inscrivent pas forcément dans la structure syntaxique recherchée. Cela occasionne beaucoup de bruit et fait baisser la précision. La moyenne harmonique entre les scores de précision et de rappel est obtenue en calculant la F-mesure suivant l'Équation 3. Les mesures de précision et de rappel varient en sens inverse si bien qu'une méthode tendant à améliorer la précision tend à dégrader le rappel.

$$\text{Précision} = \frac{\text{RestPert}}{\text{Rest}}$$

**Équation 1.** Calcul du score de précision

$$\text{Rappel} = \frac{\text{RestPert}}{\text{Pert}}$$

**Équation 2.** Calcul du score de rappel

$$\text{F - mesure} = \frac{(2 * \text{précision} * \text{rappel})}{(\text{précision} + \text{rappel})}$$

**Équation 3.** Calcul du score de F-mesure

C'est sur la base de ces trois scores que nous avons évalué la représentativité des contextes syntaxiques extraits.



#### 4. Résultats et discussion

En termes de résultats, le déroulement du cadre méthodologique de l'étude a permis de mettre en relief deux types de contextes syntaxiques, qui sont caractéristiques de l'insertion des emprunts du français aux langues burkinabè dans *Rougbêinga* de Norbert ZONGO.

##### 4.1. Le contexte définitoire

Ce contexte favorise la transmission du message véhiculé en donnant une première définition à l'emprunt. Dans l'extrait de *Rougbêinga* (p. 99), « La mienne se traduit par mon seul nom (Nanga **signifie** scorpion) », le marqueur de glose *signifie* permet de définir ou de traduire en français l'emprunt au moré *Nanga*. De plus, les parenthèses y sont utilisées pour mettre en relief la glose définitoire. Les résultats montrent que les emprunts insérés dans un contexte définitoire sont exclusivement des lexies empruntées au moré comme nous pouvons le voir dans le Tableau 1. Il s'agit de lexies empruntées pour dénommer une réalité dont la description est faite en français. Les emprunts insérés dans ce contexte sont donc suivis d'une définition permettant au lecteur de se faire une idée de la réalité désignée. Concernant ce contexte, le modèle syntaxique utilisé pour insérer les emprunts répond à la structure « Emprunt +, + glose définitoire ». L'insertion d'emprunts dans ce contexte syntaxique dénote d'un souci de favoriser la transmission du message véhiculé.

Patron syntaxique	Lexie empruntée	Langue d'emprunt	Contexte
[frpos="NOM"][frpos="PUN:cit"]?[frpos="PUN"] [word=cet   cette"][] []	<i>wakiiré</i>	moré	" <b>wakiré</b> ", cette fameuse pièce de cinq francs pour les impôts (p. 9)
	<i>Pouirga</i>	moré	pour vous apprendre que je suis le " <b>Pouirga</b> ", cet oiseau qui vole le jour avec les autres oiseaux et dort dans les trous avec les rats la nuit venue. (p. 136)
	<i>nobéré</i>	moré	c'est le <b>nobéré</b> , cette chenille au cocon très dur qui blesse mortellement tous ceux qui marchent sur elle et qui se métamorphose en un très joli papillon (p.26)
[frpos="NAM"][frpos="PUN"] [frpos="NOM   PRO:DEM"][] []	<i>Tygary</i>	moré	Nous avons le grand <b>Tygary</b> , fétiche qui ne craint que le grand Dieu seulement. (p. 100)
	<i>Lâba</i>	moré	<b>Lâba</b> , cet oiseau qui joue chaque fois au blessé pour égarer ses poursuivants.
[word="nanga"%cd] [word="signifie"]	<i>Nanga</i>	moré	La mienne se traduit par mon seul nom ( <b>Nanga signifie scorpion</b> ). (p. 97)

Tableau 1. Emprunts insérés dans un contexte définitoire

#### 4.2. Le contexte de dénomination

L'insertion des emprunts dans un tel contexte est généralement marquée par la traduction d'un objet ou d'une réalité existant ou non dans l'univers du français en une langue burkinabè. Dans ce contexte, la lexie empruntée est mise en avant suivie de sa traduction en français, comme nous pouvons le voir dans le Tableau 2. Différentes structures syntaxiques interviennent dans la description du contexte de dénomination. Ainsi, dans la structure « Emprunt + déterminant + nom », la traduction se positionne comme une expansion du nom, car l'élément typographique séparant l'emprunt de sa traduction est l'espace. Une telle structure syntaxique vise à mettre la langue burkinabè et le français sur le même pied dans l'écriture romanesque. Par ailleurs, l'insertion d'emprunts dans le contexte de dénomination est un moyen utilisé par l'auteur de l'œuvre pour prendre en compte des réalités socio-culturelles burkinabè dans la narration de l'histoire.

Patron syntaxique	Lexie empruntée	Langue d'emprunt	Contexte
[frpos="NOM   NAM"] [frpos="DET:ART"] [frpos="NOM   ADJ"]	<i>Mouhoun</i>	bwamou	Seul <b>Mouhoun le fleuve</b> donnait encore signe de vie. (p. 7)
	<i>Namuny</i>	bwamou	<b>Namuny la hyène</b> , dans les collines, entamait la mélodie de ses ancêtres (p. 8)
	<i>Saaga</i>	moré	<b>Saaga la pluie</b> s'était évanouie dans les hauteurs (p. 7)
	<i>Wounhoun</i>	bwamou	<b>Wounhoun la rivière</b> n'était plus qu'un large, long et difforme hieroglyphe. (p. 7)
	<i>Zouanga</i>	moré	certain éléments tentent de vivre comme <b>Zouri la chenille</b> (p. 14)
	<i>Zouri</i>	moré	certain éléments tentent de vivre comme <b>Zouri la chenille</b> ou <b>Zouanga la mouche</b> (p. 14)
	<i>Wongo</i>	moré	sans lequel la bête envahit l'homme, comme <b>Wongo la mauvaise herbe</b> envahit le champ de mil. (p. 8)
[frpos="NOM"] [word="appelé"]	<i>Zirgou</i>	moré	Un <b>zirgou appelé tirain</b> (p. 88)
[word="appelle.*"] [frpos="ADV"] [frpos="PUN:cit"] []	<i>Touk-touèga</i>	moré	On l' <b>appelle</b> aussi « <b>Touk-touèga</b> » (p. 96)
[frpos="NOM"] [frpos="PUN"] [frpos="NOM	<i>lôhô</i>	bwamou	Houndé le <b>lôhô (village)</b> des braves guerriers (p. 7)

Patron syntaxique	Lexie empruntée	Langue d'emprunt	Contexte
ADJ"%cd][?][?][frpos="PUN"]	<i>timbowni</i>	bwamou	le <b>timbowni (grand tam-tam)</b> grondait (p. 8)
	<i>sagbo</i>	moré	comme s'ils avaient avalé une grosse tartine de <b>sagbo (gâteau de mil)</b> . (p.52)
	<i>kounga</i>	moré	Faites sonner le <b>kounga (tam-tam de guerre)</b> . (p. 104)

Tableau 2. Emprunts insérés dans un contexte de dénomination

### 4.3. Représentativité des contextes

Dans *Rougbêinga*, les noms communs empruntés aux langues burkinabè bwamou et moré mettent un point d'honneur sur le contexte de dénomination par rapport au contexte définitoire. En effet, le contexte de dénomination présente une F-mesure plus élevée (0,73) par rapport au contexte définitoire (0,67) comme cela est visible dans le Tableau 3. En termes de précision, les patrons conçus pour l'extraction du contexte de dénomination présentent une valeur plus élevée (0,81). Une telle représentativité laisse percevoir différentes interprétations. Dans la perspective de la lutte anticoloniale dans laquelle s'inscrit le roman, la dominance du contexte de dénomination atteste d'un besoin de valorisation de la culture burkinabè, voire africaine, violée et violentée par le colon. Elle témoigne également que la dénomination constitue un facteur essentiel de véhicule des réalités socio-culturelles burkinabè dans le roman. Par ailleurs, en insérant dans un texte écrit en français des lexies empruntées à des langues burkinabè, le romancier Norbert ZONGO établit un rapport d'équivalence entre ces langues dans l'écriture romanesque et affirme une volonté de réhabilitation de la culture burkinabè dans son ensemble.

	Précision	Rappel	F-Mesure
Contexte définitoire	0,67	0,67	<b>0,67</b>
Contexte de dénomination	0,81	0,67	<b>0,73</b>

Tableau 3. Représentativité des contextes étudiés

### Conclusion

Il ressort de cette étude que la méthode Extranéo peut servir pour une extraction des contextes des emprunts du français aux langues burkinabè. Son application à *Rougbêinga* de Norbert ZONGO a permis non seulement d'extraire les emprunts et leurs contextes, mais aussi de mesurer la représentativité des contextes syntaxiques extraits. L'insertion des emprunts dans ces contextes traduit une double réalité. Il s'agit non seulement de marquer l'espace littéraire burkinabè d'une singularité, mais aussi de véhiculer un message chargé de valeurs socio-culturelles burkinabè. En cela, *Rougbêinga* de Norbert ZONGO se positionne comme un objet de valorisation de la culture burkinabè en ce sens que la langue est le véhicule par excellence de la culture, dominée soit-elle. Dès lors, cette étude ouvre des perspectives pour une activité lexicologique portant sur les emprunts du français aux

langues burkinabè dans l'espace littéraire burkinabè à partir d'une approche automatique permettant de prendre en compte différentes œuvres littéraires burkinabè dans une perspective comparative.

### Références bibliographiques

- Bloomfield, L. (1935). *Language*, Allen and Unwin, London.
- Cartier, E. (2011). Utilisation des contextes dans le cadre dictionnaire : état des lieux, typologie des contextes, exemple des contextes définitoires ». *Actes des huitièmes Journées scientifiques du Réseau de chercheurs Lexicologie, terminologie, traduction*, Lisbonne, 15-17 octobre 2009, 619-632.
- Cartier, E. (2019). Emprunts en français contemporain : étude linguistique et statistique à partir de la plateforme Néoveille. *L'emprunt en question (s) : conceptions, réceptions, traitements lexicographiques*, Limoges : Lambert-Lucas, 1-26.
- Cartier, E. & Sablayrolles, J.-F. (2009). Néologismes, dictionnaire et informatique. *Cahiers de lexicologie*, Centre National de la Recherche Scientifique, (93), 175-192
- Daïla, B. M. (2017). *Le français ma houe, mon travail au Burkina Faso*, Éditions universitaires européennes.
- Evert, S. & Hardie, A. (2011). Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium». *Proceedings of the Corpus Linguistics 2011 conference*, Birmingham, UK., [http://www.birmingham.ac.uk/documents/college\\_artslaw/corpus/conference-archives/2011/Paper-153.pdf](http://www.birmingham.ac.uk/documents/college_artslaw/corpus/conference-archives/2011/Paper-153.pdf), 1-21.
- Dubois, J. et al. (2001). *Dictionnaire de linguistique*, Larousse, Paris.
- Grevisse, M. & Goosse, A. (2011). *Le bon usage*, 15<sup>e</sup> édition, De Boeck & Duculot, Paris.
- Gross, G. (2010). Sur la notion de contexte. *Meta*, (551), 187-197.
- Harris Z. S. (1968). *Mathematical Structures of Language*, John Wiley & Sons, New-York.
- Heiden, S. & al. (2010). TXM : Une plateforme logicielle open-source pour la textométrie-conception et développement ». *Processings of 10th International Conference on the Statistical Analysis of Textual Data*, (2), 1021-1032.
- Kedrebéogo, G. & al. 1988). *Burkina Faso : Carte linguistique*, CNRST, Ouagadougou.
- Kéïta, A. & Pardevan, E. (2020). Le traitement des emprunts du mooré et du dioula au français dans les écoles primaires bilingues du Burkina Faso. *Akofena*, (001), 93-112.
- Kent, A. & al. (1955). Machine literature searching VIII. Operational criteria for designing information retrieval systems. *American Documentation*, 6, (2), 93-101.
- Koama, C. (2004). Analyse des créations lexicales et emprunts dans le Journal du jeudi. Mémoire de Maîtrise, Université de Ouagadougou, U.F.R. LAC, en Lettres modernes.
- Loubier, C. (2011). De l'usage de l'emprunt linguistique, Office québécois de la langue française, Montréal.
- Mela, A. & al. (2011). Mixer les moyens pour extraire les gloses. *Actes de la conférence EGC : Extraction et Gestion des Connaissances*, Janvier 2011, Brest, France, 95-106.
- Mortureux, M.-F. (1984,). La dénomination, approche socio-linguistique. *Langages*, (76), 95-112.

- Natama, T. J. (2021). Analyse sémantique de mots moore dans *Rougbêinga* de Norbert ZONGO. *DELLA/AFRIQUE*, numéro Hors série, spécial covid-19, 272-289.
- Ouédraogo, Y. (2000). Le français basilectal dans la littérature burkinabè. *Actes de la XVIIIe biennale de la langue française*, tenue à Ouagadougou sur le thème « L'expression du droit. Le français, langue africaine et internationale. La jurisfrancité. Le Burkina Faso et la francophonie », Paris, 381-390.
- Pincemin, B. & al. (2006). Concordanciers : Thèmes et variations. *Actes des 8<sup>es</sup> Journées internationales d'Analyse statistique des Données Textuelles (JADT 2006)*, Jean-Marie VIPREY et al. (éds), Besançon : Presses Universitaires de Franche-Comté, (II), 773-784.
- Pinel-sauvagnat, K. & Mothe, J. (2013). Mesure de la qualité des systèmes de recherche d'information. *Ingénierie des systèmes d'information*, (10), DOI :10.3166/ISI.X.X.1-27, 1-27.
- Poibeau, T. et Nazarenko, A. (1999). L'extraction d'information, une nouvelle conception de la compréhension de texte ? *Traitement Automatique des Langues*, 40 (2), pp. 87-115.
- Prignitz, G. (1994). Rôle de l'argot dans la variation et l'appropriation : le cas du français au Burkina. *Langue française*, (104), 49-63.
- Rijsbergen, V. (1979). *Information Retrieval* (second edition), London, Butterworths.
- Rondeau, G. (1986). Introduction à la terminologie, Gaétan Morin, Chicoutimi.
- Sablayrolles, J.-F. (2017). Les néologismes. Créer des mots nouveaux aujourd'hui, Garnier, Paris.
- Schmid, H. (1995). Treetagger, a language independent part-of-speech tagger. *Institut für Maschinelle Sprachverarbeitung Universität Stuttgart*, (43), 1-28.
- Zongo, N. (2009). *Rougbêinga*, Harmattan Burkina, Ouagadougou.
- Zoumbara, C. et al. (2020). Combinaison d'étiqueteurs morpho-syntaxiques, de lexiques flexionnels et de marqueurs de glose pour extraire les néologismes en français du Burkina. *Actes des 15es Journées internationales d'analyse statistique des données textuelles (JADT)*, 16-19 juin 2020, Toulouse, France.

**Annexe : Jeu d'étiquettes de l'étiqueteur morphosyntaxique TreeTagger**

Étiquette	Signification
ABR	Abréviation
ADJ	Adjectif qualificatif
ADV	Adverbe
DET:ART	Article
DET:POS	déterminant possessif
INT	Interjection
KON	Conjonction
NAM	nom propre
NOM	nom commun
NUM	Nombre
PRO	Pronom
PRO:DEM	pronom démonstratif
PRO:IND	pronom indéfini
PRO:PER	pronom personnel

Étiquette	Signification
PRO:POS	pronom possessif
PRO:REL	pronom relatif
PRP	Préposition
PRP:det	préposition plus article
PUN	Ponctuation
PUN:cit	Citation
SENT	fin de phrase
SYM	Symbole
VER:cond	verbe au conditionnel
VER:futu	verbe au futur
VER:impe	verbe à l'impératif
VER:impf	verbe à l'imparfait
VER:infi	verbe à l'infinitif
VER:pper	participe passé d'un verbe
VER:ppre	participe présent d'un verbe
VER:pres	verbe au présent
VER:simp	verbe au passé simple
VER:subi	verbe à l'imparfait du subjonctif
VER:subp	verbe au présent du subjonctif