

Collocations dans les sciences du langage : enjeux théoriques et applicatifs

Collocations in language sciences: theoretical and applicational issues

Aimeline RASOANANTENAINA¹ 
Université d'Antsiranana / Madagascar
rn.aimeline@gmail.com

Cécile Marie-Ange MANOROHANTA² 
Université d'Antsiranana / Madagascar
cmanoroh@gmail.com

Reçu: 30/04/2024,

Accepté: 06/06/2024,

Publié: 30/06/2024

Résumé

Les collocations, étant marquées jusque-là par l'abondance de descriptions ponctuelles, font l'objet de plusieurs débats qui rappellent la nécessité d'une mise au point méthodologique et terminologique. Pourtant, malgré cet intérêt, les contours du phénomène restent vagues. Il reçoit diverses interprétations et la divergence de définition a fait des collocations une question complexe et difficile à aborder. Dans les sciences du langage, les débats sont centrés autour du problème de la modélisation linguistique du phénomène et sur les impacts des modèles proposés dans plusieurs domaines applicatifs.

Mots-clés : combinatoire lexicale - cooccurrence - linguistique appliquée - modélisation,

Abstract

Collocations, being marked until now by the abundance of specific descriptions, are the subject of several debates which recall the need for methodological and terminological clarification. However, despite this interest, the contours of the phenomenon remain vague. It receives various interpretations and the divergence in definition has made collocations a complex and difficult issue to address. In the language sciences, debates are centered around the problem of linguistic modeling of the phenomenon and on the impacts of the models proposed in several application domains.

keyword : applied linguistics - co-occurrence - lexical combinatorics - modelling,

* Auteur correspondant : Aimeline RASOANANTENAINA

Introduction

De par l'universalité du phénomène collocationnel, son attribut quantitatif très important (Dechamps, 2004) ainsi que le flottement terminologique qui l'entoure, l'intérêt de la prise en compte des collocations est aujourd'hui devenu une évidence. Dans les divers écrits qui traitent des phénomènes collocationnels, la « collocation » a reçu différentes dénominations, entre autres « groupement usuel » (Bally, 1909), « combinaisons phraséologiques » (Blumenthal et Hausmann, 2006), « phrasème sémantique », « semi-phrasème » (Mel'čuk, 2003). Toutefois,

« collocation » est le terme qui a été majoritairement accepté pour définir les cooccurrences lexicales restreintes entre deux ou plusieurs constituants. Nous avons préféré garder cette appellation pour ne pas multiplier les termes et entraîner la confusion.

Ce travail, qui met l'accent sur les problématiques de définition et d'ordre théorique, abordera les points suivants : exposer les discussions autour des différents enjeux que soulève la modélisation des collocations, en partant des définitions présentées dans les traditions linguistiques, ce qui nous amènera à montrer les diverses applications possibles tels que dans la lexicologie/lexicographie, la traduction et la didactique des langues, faisant ainsi apparaître les avancés et les contraintes des méthodes analysées.

1. Collocations dans les sciences du langage

Bien que les collocations aient été longuement étudiées, la façon de les concevoir ne fait pas l'unanimité. Dans la tradition linguistique, nombreuses confusions sont attachées à ce phénomène lexical ; certains le classe parmi les combinatoires libres et d'autres dans les syntagmes figés, créant ainsi un flottement terminologique (Verlinde, Binon et Selva, 87 :2006). Avec la diversité des critères pour le décrire et le définir, le phénomène collocationnel est difficile à formaliser, aussi les linguistes du 19^e siècle avaient-ils tendance à l'ignorer (Blumenthal et Hausmann, 7 :2006), de même que les grammairiens et ceux qui travaillent sur la sémantique lexicale (Williams, 2003).

1.1. Enjeux théoriques et applicatifs de la modélisation

Pourquoi travailler sur les collocations ? Dans un contexte actuel de recherche d'informations, parmi les choses qui préoccupent la société, il y a la recherche des discours propres à un domaine, propres à un auteur ou encore à un personnage politique. Gledhill (1997) considère ces cooccurrences comme le témoin de la présence d'un discours et jouent, par conséquent, un rôle dans la définition du « genre » textuel. Les terminologues s'intéressent beaucoup aux comportements de cette combinatoire lexicale dans les textes spécialisés.

Dans le domaine de l'informatique, l'exemple de « créer un fichier » montre qu'il y a une préférence pour certaines unités lexicales au détriment d'autres (L'Homme, 90 :2003). Ainsi « fichier » se combine de préférence avec « créer » plutôt qu'avec d'autres verbes qui véhiculent plus ou moins le même sens :

- Créer (N₀hum, fichier)
- *Etablir (N₀hum, fichier)
- *Concevoir (N₀hum, fichier)

Le non-spécialiste qui sera appelé à rédiger ou à faire la traduction des textes dans un domaine spécifique ne maîtrise pas toujours ces usages lexicaux ainsi que les compatibilités entre ces unités lexicales.

Les recherches sur les cooccurrences discursives répondent donc aujourd'hui à différents enjeux : politique, économique, scientifique, culturel, etc. Les collocations relèvent donc de l'idiosyncrasie puisqu'elles marquent les particularités linguistiques propres à un individu

(permettant par exemple de définir le style d'un auteur ou encore d'un personnage politique), spécifiques à un certain type de discours.

Le flou terminologique qui entoure les descriptions linguistiques présente également des impacts au niveau de la didactique des langues. Certains auteurs avancent que pour pouvoir maîtriser parfaitement une langue, l'acquisition de ses collocations est une condition sine qua non. Ce phénomène a été admis comme étant central dans la compétence linguistique des locuteurs natifs (Williams, 2003). On note cependant que certains auteurs ont ouvert le champ à l'analyse et à l'enseignement des collocations dans le cadre du Français sur Objectif Universitaire (FOU) et du Français sur Objectif Spécifique (FOS). Cavalla (2019) fait partie des pionniers qui ont travaillé sur l'enseignement des collocations en français, sur les règles d'association et la sélection d'un collocatif possible pour une base donnée. Priesolová (2018) a mis également l'accent sur les collocations pour faire ressortir les spécificités non seulement linguistiques mais également aux niveaux socioprofessionnels.

Une autre attention pour laquelle l'identification des collocations apporte une contribution importante est l'application lexicographique. Plusieurs aspects sont à considérer : dans un dictionnaire d'encodage, quelles sont les informations linguistiques qui doivent être décrites sur l'unité lexicale ? Pour permettre un accès facile à ces informations et pour que les usagers s'en servent mieux pour améliorer leurs productions, comment devrait-on les organiser et les présenter ? Est-ce que le collocatif est une entité lexicale suffisamment autonome sémantiquement pour pouvoir bénéficier d'un article lexicographique à part entière ? Alonso Ramos (53 :2003) analyse le statut et le traitement des collocatifs dans la lexicographie en proposant des critères qui permettent « *de décider si la création d'un article lexicographique pour un collocatif donné est nécessaire.* ». Le double accès est devenu maintenant possible grâce aux nouvelles versions de dictionnaires électroniques. Ajouter à tout cela, ce phénomène est loin d'être marginal dans les langues puisque son importance réside dans leur omniprésence (Mel'čuk, 26 :2003). En effet, les collocations sont numériquement importantes (Grossmann et Tutin, 6 :2003) et les textes de tous les genres en regorgent.

Les différents enjeux qui viennent d'être mis en évidence confirment la nécessité de travailler sur les collocations. Se pose maintenant la question de la modélisation puisque la collocation demeure un concept qui est difficile à formaliser et n'a été, le plus souvent, étudié que par rapport à des exemples ponctuels et des cas prototypiques (Williams, 2003).

1.2. Intérêts de la modélisation

Pourquoi est-il important de modéliser les collocations et de quelle manière pourrait-on le faire ? A la première question, Polguère définit trois caractéristiques, que possèdent les collocations, qui expliquent l'intérêt que présente la modélisation :

- « 1 Elles sont universellement présentes dans toutes les langues.
- 2 Elles sont omniprésentes dans les textes, qu'ils soient oraux ou écrits.
- 3 Elles semblent plus ou moins arbitraires, ne peuvent pas se traduire mot à mot d'une langue à l'autre et sont donc très difficiles à acquérir. » (Polguère, 137 :2002)

Celle-ci peut se définir comme un mécanisme permettant de réaliser un modèle qui sera simultanément définitoire, descriptif et prédictif (Dubreil, 117 :2008). De ce fait, la modélisation

est nécessaire, « étape obligée » selon Grossmann et Tutin (8 :2003) afin de décrire, d'un point de vue théorique, ce phénomène linguistique, à partir de ses propriétés syntactico-sémantiques mais aussi d'un point de vue applicatif car elle permet de définir les caractéristiques qui permettent d'expliquer, d'anticiper ou reproduire les collocations dans divers secteurs. Mais par quels moyens ?

Après la sémantique contextuelle de Firth basée sur le critère de la co-présence des constituants, la notion de collocation suit deux cheminements distincts (Blumenthal et Hausmann, 2006) : une orientation statistique et une orientation fonctionnelle. La première acception, issue de l'approche linguistique de corpus, définit la collocation sur des bases statistiques comme étant une co-occurrence « fréquente » de lexies et par conséquent, elle est en faveur de la linguistique informatique. La seconde acception, qui décrit la collocation comme une « cooccurrence lexicale restreinte », se concentre plus sur les différentes restrictions que le nombre d'apparitions. Par exemple, on dit « pousser un hurlement » mais non « *émettre un hurlement » alors que ce verbe est parfaitement approprié au nom « son ». Cette approche sémantique (appelée également approche lexicologique-lexicographique) décrit les collocations sur des bases strictement fonctionnelles et sémantiques.

2. Le contextualisme et la linguistique de corpus

Parler du concept de collocation revient à rappeler la tradition contextualiste britannique, initiée par Firth. Pour ce précurseur, ces co-occurrences sont surtout utilisées dans une perspective textuelle. Pour ceux qui sont en faveur du contextualisme, la collocation désigne « *les relations de sens portées par des mots qui tendent à apparaître ensemble dans un contexte textuel donné* » (Dubreil, 118 : 2008). Cette définition est qualifiée de « large » parce qu'elle repose essentiellement sur des critères statistiques et c'est pourquoi, elle devient une notion essentielle dans la linguistique de corpus.

2.1.Critères statistiques

Pour les contextualistes, ce qui est important c'est d'observer les relations qui se forment dans les textes. Dans cette perspective, les collocations sont définies comme des associations de mots qui apparaissent fréquemment en co-occurrence mais qui n'entretiennent pas nécessairement une relation syntaxique directe. Pour eux, les critères sémantiques ou syntaxiques ne sont pas nécessaires et en l'absence de ces critères spécifiques, la collocation textuelle est également définie comme étant une collocation statistique. Les collocations sont alors considérées comme étant des « *unités polylexicales codées en langue* » (Blumenthal et Hausmann, 4 :2006) et de ce fait, comme les autres éléments du lexique, étant étroitement liées aux notions de fréquence et de l'usage, elles sont mémorisées par les utilisateurs et présentent souvent une unité référentielle puisqu'elles sont associées à un contenu sémantique prédéterminé.

La notion de co-occurrence habituelle est le premier critère présenté pour définir les collocations. Il est difficile de définir ce qu'on entend par « habituel ». La notion de fréquence, bien qu'elle soit importante, reste floue car ce critère ne permet ni de décrire les collocations ni de les extraire à partir d'une application informatique. Ainsi, pour être plus explicite et opératoire, les critères de type statistique fondés sur les notions de fréquence et d'usage ont été introduits en insistant sur l'utilisation de grand corpus pour l'étude linguistique. Parmi les

premiers travaux informatisés sur la collocation, on peut citer celui de Rogue qui a élaboré une théorie distributionnelle dans les années 70 pour rendre explicite la notion de collocation en termes statistiques. Néanmoins, chaque outil statistique utilisé semble n'être approprié qu'à une application bien distinctive, ce qui implique que la mesure statistique mobilisée ne peut être applicable que pour une application précise. Les outils statistiques présentent donc des limites non seulement d'ordre pratique mais aussi théorique, comme le montre le cas de la coordination elliptique de structure Adjectif Adjectif en malgache que nous allons examiner.

2.2.Limites des mesures statistiques : cas de la coordination elliptique

Une co-occurrence, un « segment répété » selon la terminologie de Mejri (196 :2008), peut ne pas être une collocation. Concevoir la collocation comme une fréquence de proximité entre au moins deux éléments dans les énoncés n'est pas nécessairement pertinente du point de vue linguistique puisque toute co-occurrence ne correspond pas obligatoirement à une collocation.

C'est le cas par exemple de la coordination elliptique dans les constructions adjectivales malgaches. En effet, les termes de coordination peuvent différer pour certaines constructions adjectivales. Malgré cet effacement au niveau phonologique, ils sont néanmoins présents au niveau sémantique. Dans les exemples suivants, nous allons nous intéresser au phénomène de l'asyndète qui est une ellipse qui supprime les termes de liaison logiques. Il est question d'observer les propriétés particulières des constructions elliptiques Adjectif Adjectif dans une perspective transformationnelle :

Mahia lava Dolin.

Mince grand Dolin

= *Mahia (E + no sady) lava Dolin.*

(E + Sady) Mahia (E + no) lava Dolin.

(E + Aussi) mince (E+que) grand Dolin

Dolin est aussi mince que grand.

Dolin est mince et grand.

Les conjonctions de coordination servent à relier des éléments de même rang syntaxique. Les deux adjectifs composant la construction adjectivale appartiennent à la même classe, celle de l'apparence physique. La coordination met en jeu deux structures sémantiques parallèles qui ne sont que partiellement réalisées d'un point de vue syntaxique :

Söla vantony baban'i Solo.

Söla (E + kanêfa mbola) vantony baban'i Solo.

Chauve (E + mais encore) jeune père de Solo.

Le père de Solo est encore jeune mais chauve.

Manta lo katakata aby irô.

(E + Mbola) manta (E + kanêfa) (E + êfa) lo (E + tapany) katakata aby irô.

(E + Encore) vertes (E + mais) (E + déjà) pourries (E + en partie) ces bananes.

Ces bananes sont encore vertes mais elles sont déjà pourries en partie.

L'ellipse, dans ces exemples, consiste à retracer des éléments nécessaires à la construction, mais non au sens. Ce qui nous a permis de confirmer qu'il n'est pas question de collocations à adjectifs prédicatifs dans ces co-occurrences mais plutôt de constructions elliptiques.

On est tenté d'avancer, comme Williams (2003), que les résultats fournis par les mesures statistiques forment des collocations « candidats » et qu'il faudra d'autres critères pour confirmer que la relation entre les composants est linguistiquement fondée. Du point de vue du traitement automatique, les études statistiques ne sont pas suffisantes pour étudier les collocations car elles mettent en évidence uniquement des cooccurrences. Il faut compléter les mesures statistiques par des analyses linguistiques pour établir quelles cooccurrences sont des collocations ou inversement, les hypothèses linguistiques devraient être formulées à partir des corpus. Suivant cet axe, Blumenthal allie l'importance du corpus et l'intérêt linguistique dans son « profil combinatoire » en appliquant la définition restreinte (base - collocatif) dans son outil statistique.

2.3. Le profil combinatoire

La définition de la collocation dans l'approche contextualiste varie selon les problématiques posées et l'outil statistique adopté. Par ailleurs, une cooccurrence n'est pas nécessairement une collocation : elle peut être un segment répété, une figure stylistique, etc. d'où l'importance des critères linguistiques. Plus récemment, les linguistes et les informaticiens admettent la nécessité de consulter des corpus, donc des notions contextualistes, mais le tout est de savoir à quelle finalité on va utiliser l'informatique.

Blumenthal (2011) a élaboré un outil appelé « profil combinatoire » qui permet d'identifier les collocations dans un corpus discursif à partir de la fréquence d'apparition de l'association lexicale d'une part et l'inventaire des classes de bases dans la description des collocatifs d'autre part, dont les résultats « *peu[vent] se mettre au service de la synonymie, de l'apprentissage systématique du vocabulaire, de la traduction automatique ou d'autres activités.* » (Blumenthal et Hausmann, 5 :2006).

En conclusion, les deux acceptions du terme collocation ne s'opposent pas radicalement car le profil combinatoire permet d'avoir une plus large possibilité d'observations.

3. La tradition continentale et les descriptions binaires des collocations

La seconde tradition, continentale, principalement marquée par Bally, Mel'čuk, Cowie, Hausmann et, plus récemment, Binon, Verlinde, Grossmann, Tutin et autres s'est développée dans le contexte de la lexicologie et la lexicographie, en adoptant une définition « restreinte » fondée sur des paramètres sémantiques où la notion de binarité est centrale. Les collocations sont alors définies comme une séquence composée d'une base et d'un collocatif :

« La base est un mot (plus précisément l'acception d'un mot, appelée aussi « lexie ») que le locuteur choisit librement parce qu'il est définissable, traduisible et apprenable sans le collocatif. Le collocatif est un mot (ou l'acception d'un mot) que le locuteur sélectionne en fonction de la base parce qu'il n'est pas définissable, traduisible ou apprenable sans la base. » (Blumenthal et Hausmann, 4 :2006)

La binarité se présente sous deux aspects dans les critères définitoires de la collocation. Nous allons examiner dans les parties qui vont suivre les différentes propriétés que recouvre cette description binaire ainsi que les modèles de description qui ont été adoptés.

3.1. Binarité catégorielle

Certaines études partagent le fait que la collocation soit marquée par une « bipartition » où l'association est composée principalement de deux mots ou de deux lexies (Blumenthal et Hausmann : 4 : 2006) comme dans les exemples suivants :

- « Grièvement (collocatif) blessé (base) »
- « Une bouffée de (collocatif) colère (base) »
- « Un célibataire (base) endurci (collocatif) »

Ce caractère duel n'est pas aussi marqué dans d'autres recherches, comme celle de Siepmann qui avance l'existence des collocations « tripartites » (102 :2006). Il paraît plus évident en effet de parler d'« association de constituants » plutôt que de lexies dans de nombreux cas puisque les collocations englobent davantage que deux termes, comme le cas des « collocatifs à valeur métaphorique » (Polguère et Mel'čuk, 70 :2006) que l'on retrouve dans « méchant comme une teigne ». Les lexicographes semblent considérer le caractère binaire suffisamment opérationnel puisque plusieurs ouvrages lexicographiques ont été vulgarisés suivant cette définition, comme le Dictionnaire d'apprentissage du français langue étrangère (Dafles) de Verlinde, Selva et Binon (2003) et le Lexique Actif du Français de Polguère (2000).

Si le caractère binaire est institutionnalisé dans les différents dictionnaires, qu'ils soient destinés au grand public ou à des publics d'apprenants, certains linguistes comme Bartsch (2004) ou Siepmann (2006) le contestent parce-que certaines expressions lexicalisées et compositionnelles dépassent l'association binaire (binarité catégorielle). A cette remarque, Tutin (52 - 54 :2013) explique que de nombreuses collocations se combinent par superposition ou enchaînement.

Pour le premier cas, elle a pris l'exemple de « avoir une peur bleue » qui résulte de la superposition de deux collocations indépendantes « avoir peur » et « peur bleue » partageant une base commune, le nom « peur ». Le second cas, appelé également « collocation récursive », regroupe les cas où la base ou le collocatif peuvent être eux-mêmes des collocations. Ainsi, si on examine de plus près la séquence « tomber dans l'oubli », on s'aperçoit qu'elle résulte de la combinaison entre le collocatif « tomber », qui est un verbe support exprimant une valeur inchoative, et d'une base « dans l'oubli », qui est elle-même une collocation composé d'un collocatif « dans » exprimant l'état et d'une base nominale « oubli ». Aussi, Tutin (50 :2013) reformule-t-elle le critère de binarité de manière beaucoup plus englobante en insistant sur le fait que la collocation met en jeu une relation syntaxique entre des constituants qui peuvent être une unité lexicale simple ou complexe.

3.2. Binarité fonctionnelle

Le deuxième aspect de la binarité, dite fonctionnelle, est marqué par la hiérarchie base-collocatif qui sous-entend un statut différent entre les deux composants : la base est dite sémantiquement

autonome tandis que le collocatif ne l'est pas (entre autres, Blumenthal 2011 ; Grossmann et Tutin 2003 ; Alonso Ramos 2003). Mel'čuk (27 :2003) insiste d'ailleurs sur ce critère, qu'il qualifie de trait définitoire, car le collocatif a une dépendance fonctionnelle vis-à-vis de la base. Suivant ce point de vue, c'est cette dernière, autonome et choisit librement par le locuteur, qui contrôle et sélectionne le collocatif en fonction d'un sens spécifique que l'on veut formuler et non l'inverse. Cette hiérarchie qui existe entre les éléments lexicaux qui composent la collocation fait qu'on identifie toujours une base (peur, célibataire, gifle, examen) et un collocatif (bleue, endurci, administrer, passer).

Cependant, pour certains groupes de collocations, la binarité base-collocatif n'est pas toujours facile à reconnaître. Autrement dit, il ne serait pas toujours possible de faire la distinction entre base et collocatif parce qu'il ne semble pas toujours possible de mettre en évidence l'élément dominant parmi les composants de l'association. Selon Tutin (55 :2013), si certaines hésitations sont possibles pour certaines séquences, c'est parce-que certains éléments sont désémantisés et présentent une ambiguïté sémantique et syntaxique. Pour l'expression « une meute de loup » par exemple, deux interprétations sémantiques (et syntaxiques) sont possibles : quantitative et qualitative. Si l'on choisit l'interprétation quantitative, la base est l'élément « loup » et le collocatif, l'élément quantitatif « une meute de ». En revanche, si on met l'accent sur l'interprétation non quantitative la base est alors « meute » et le collocatif « loup » est analysé comme étant un complément de nom régis. Cette hiérarchie base-collocatif est basée sur des critères sémantiques puisque ces associations syntagmatiques restreintes sont présentées comme des suites préfabriquées à situer dans le continuum des expressions polylexicales.

D'ailleurs, Tutin et Grossmann (2002) proposent, à partir de l'idiomaticité du collocatif et le critère de l'opacité, une typologie sémantique des collocations. Ainsi, trois grandes classes ont été répertoriées. Les collocations sont dites opaques lorsque le sens de la séquence est non transparent. Du point de vue sémantique, l'adjectif « bleue » est arbitraire, donc imprédictible car il n'a plus le même sens en dehors de l'association « une peur bleue ». Les collocations transparentes, quant-à-elles, regroupent des associations où le sens du collocatif est facilement décodé et interprétable en cooccurrence avec la base, c'est le cas de « grièvement » dans « grièvement blessé ». Enfin, la dernière classe, les collocations qualifiées de « régulières », quand le collocatif est approprié à une base unique puisque le sens de celle-ci est intégré dans sa définition, que l'on retrouve par exemple dans « nez aquilin ».

Lorsque l'on cherche à décrire le lexique, des choix théoriques et descriptifs doivent être déterminés. Polguère (2002) considère les fonctions lexicales comme une réponse adéquate à la modélisation des collocations. Cependant, bien que les fonctions lexicales constituent un apport fondamental dans la description des collocations, la description syntaxique proposée est un peu moins satisfaisante (Grossmann et Tutin, 13 :2003). Tutin (2013) suggère que pour améliorer la modélisation des collocations, il faudrait un traitement beaucoup plus méthodique des propriétés syntaxiques, en insistant sur les particularités des collocatifs qui régissent la motivation sémantique. Sur le plan sémantique, ces modèles linguistiques répondent à plusieurs types de régularités qui s'organisent autour de deux sources de motivation : les régularités sémantiques d'un côté et les procédés sémantiques productifs de l'autre.

3.3. Les régularités sémantiques et les fonctions lexicales

Mel'čuk et al. (1995) ont proposé un outil, dénommé les fonctions lexicales (FL), qui s'apparentent à des fonctions mathématiques et qui a pour objectif de modéliser non seulement les liens lexicaux mais aussi les liens de cooccurrence au sein des collocations. Le codage des collocations est caractérisé à l'aide des fonctions lexicales syntagmatiques qui permettent de systématiser les relations sémantiques qui lient les composants de la collocation. Les fonctions sont appelées lexicales puisqu'elles décrivent une relation qui existe entre une lexie et un ensemble de lexies ou d'expressions figées afin de :

« Fournir au locuteur la totalité des moyens lexicaux nécessaires à l'expression la plus riche, la plus variée et la plus complète de la pensée et, en même temps, de garantir le choix le plus précis de la formulation appropriée. » (Mel'čuk et al., 129 :1995)

Cette relation repose sur le critère d'universalité, on a alors les FL standard et celles qui sont non-standard. Les premières appellent le critère évoqué ci-dessus car elles possèdent cette propriété au sens où il est possible de les appliquer à toutes les langues. Pour exprimer la fonction intensive par exemple, la base « faim » en français, et en malgache *mafo*, ont recours à un collocatif identique « mort de » traduit par *matin'ny* en malgache dans les deux langues :

Matin'ny mafo (malgache)
Mort de faim (français)

Pour l'expression de la même fonction cependant, des différences peuvent être constatées d'une langue à l'autre. Pour intensifier le mot-clé « pluie » (« rain » en anglais) par exemple, la langue française et la langue anglaise sélectionne de valeurs distinctes (Jousse, 2010). Pour la première, la composante d'intensité porte sur le débit alors que pour l'anglais, elle porte sur le poids :

Forte pluie (intensité : le débit de la pluie)
Heavy rain (intensité : le poids de la pluie)

Une fonction possède donc plusieurs éléments comme valeurs. Ces différences de valeurs, exprimées par les collocatifs, sont fortement liées aux connotations culturelles mais le sens de la fonction n'en reste pas moins stable.

La standardité des fonctions lexicales présente une limite dans la mesure où certaines fonctions perdent une partie des caractéristiques qui définissent les fonctions standard : il s'agit des fonctions non standard. Ces dernières, contrairement aux fonctions standard, ne peuvent prétendre appartenir au cadre de l'universalité. A ce titre, elles n'apparaissent pas en principe dans toutes les langues. Une fonction lexicale non-standard ne s'applique qu'à un très petit nombre de lexies (peu de mot clés, voire à un seul). Ce caractère spécifique fait qu'elle possède peu de valeurs (voir une seule), c'est le cas de « café noir » où le collocatif adjectif « noir » signifie « sans produit laitier ». Puisque les FL non standard sont imprévisibles, on ne peut pas les inventorier de façon systématique et les décrire avec un formalisme prédéterminé. L'unique manière de les recenser consiste à les retrouver dans les textes et définir chacun de ses sens isolément.

La description à partir de l'étiquetage d'un lien qui unit le collocatif à la base au moyen d'une FL permet de décrire de façon formelle la relation sémantique et syntaxique qui les réunissent, aboutissant ainsi au classement des collocations. Dans ce sens, le modèle satisfait aux exigences prescrites par le principe d'une modélisation de collocations. En revanche, les fonctions lexicales non standard constituent un obstacle sérieux, de par leurs natures spécifiques et non systématique, on ne peut les prédire et de ce fait, elles ne peuvent pas être caractérisées et recensées de façon méthodique. Une des raisons principales de ces difficultés, selon Grossmann et Tutin (17 :2003), est qu'en vue de modéliser les collocations, les auteurs ont priorisé une « approche purement lexémique des collocations » dans la mesure où ils considèrent les collocations comme des « cooccurrences spécifiques » où le collocatif est associé à une seule et unique base alors que le plus souvent, on se retrouve à des « moules syntactico-sémantiques ». Le caractère trop limitatif de cette définition du phénomène collocationnel pose problème étant donné que la base peut être remplacée par des synonymes : un étudiant peut rater (un test + un examen + un concours), etc.

3.4. Les procédés sémantiques productifs

La question des procédés sémantiques productifs n'a pas été abordée dans le modèle lexical des FL. Il est question d'analyser les principaux mécanismes sémantiques permettant de générer les collocations. Il s'agit notamment de préciser le choix non aléatoire des règles d'association qui réunit la base et le collocatif en fonction des contraintes sémantiques. Les études sur les collocations lexicales font état de régularités et ce, à plusieurs niveaux : dans les schémas syntaxiques employés, dans les réalisations lexicales associant des paradigmes sémantiques et dans les relations sémantiques mises en jeu. Indépendamment, d'une certaine façon, du lexique employé, certaines constructions syntaxiques semblent être spécialisées dans certains traitements sémantiques (Tutin, 2013).

Les procédés sémantiques productifs concernent les collocations imagées. Les associations collocationnelles créées grâce aux procédés stylistiques (métaphore, métonymie, comparaison, etc.) sont en partie régulières et de ce fait, on peut envisager d'indiquer quelques-unes des règles qui sont à l'origine des procédés productifs. L'analyse du syntagme binominale en « de » par exemple a permis de voir qu'on peut exprimer la collection et la quantification, que l'on retrouve dans « une forêt de mâts » et « un essaim d'hélicoptères », à travers la structure N de N (Tutin et Grossmann, 2002). Dans ce contexte, le choix du nom collectif n'est pas arbitraire, il est accompagné de certains critères d'appropriation référentielle car on ne peut dire « *un essaim de mâts ». Suivant le même principe, dans « une avalanche de questions » ou « un déluge de paroles », les collocatifs exprimés par la structure Det N Prép ont une valeur quantitative acquise métaphoriquement. Il est essentiel de définir les conditions d'application référentielle liées au sémantisme des collocations pour pouvoir prédire ces associations. Ces moules productifs contribuent au caractère régulier des collocations et par conséquent, à la modélisation de ce phénomène lexical.

Conclusion

Les collocations ont été initialement définies par des considérations statistiques. Dans la tradition « continentale », on attribue une définition plus étroite de la collocation (appelée

généralement collocation lexicale restreinte) où celle-ci est décrite à partir des deux aspects de la binarité. Sur le plan sémantique, le statut des deux éléments qui composent la collocation n'est pas égal (on parle d'une asymétrie sémantique) : la base est autonome et le collocatif ajoute une caractérisation à la base sans toutefois modifier son identité sémantique. C'est autour de cette contenance sémantique que se concentre l'intérêt de cette seconde tradition. Mel'čuk y voit l'expression des fonctions lexicales qui cherchent à encoder l'apport du sens du collocatif, à partir de la base, pour exprimer une information bien précise. Dans cette optique, définir la collocation comme étant une base reliée à son collocatif se pose la difficulté de rendre compte des questions ayant trait à la motivation sémantique qui est pourtant importante en linguistique puisque même si certaines collocations sont considérées comme accidentelles, nombreuses collocations s'organisent de façon structurée autour de plusieurs dimensions linguistiques. Cependant, il reste encore la question de la polysémie et du figement qui a été peu développée dans les différentes théories sur la modélisation des collocations.

Références bibliographiques

ALONSO RAMOS Margarita (2003), « La nature des collocatifs : leur statut en tant qu'unités lexicales », in *Les Collocations : analyse et traitement*, Éd. De Werelt, pp. 46-60.

BLUMENTHAL Peter (2011), « Odeur – évolution des profils combinatoires », in *Langages* N° 181, 1/2011, pp. 53-71. URL : <https://www.cairn.info/revue-langages-2011-1-page-53.htm> (consulté le 06 novembre 2023).

BLUMENTHAL Peter et HAUSMANN Franz Josef, (2006), *Collocations, corpus, dictionnaires*, in *Langue Française* N°150, Éd. Larousse, Armand Colin. URL : https://www.persee.fr/issue/lfr_0023-8368_2006_num_150_2 (consulté le 13 septembre 2023).

CAVALLA, Cristelle, 2019, « Une entrée lexicale spécifique : les collocations ». *Palabras como puentes : estudios lexicológicos, lexicográficos, y terminológicos desde el Cono Sur*, Buena Vista Editora, pp.58-67. URL : <https://hal.science/hal-02533954/document> (consulté le 08 avril 2024).

DECHAMPS, Christina, 2004, « Enseignement/apprentissage des collocations d'une langue de spécialité à un public allophone : l'exemple de la langue juridique », *Éla. Études de linguistique appliquée* 2004/3 N°135, Klincksieck ; pp. 361 à 370. URL : <https://www.cairn.info/revue-ela-2004-3-page-361.htm> (consulté le 03 avril 2024).

DUBREIL Estelle (2008), « Proposition de typologie argumentative des collocations textuelles », in *Texte et corpus* N°3, pp. 117-126.

GLEDHILL Christopher (1997), « Les collocations et la construction du savoir scientifique », in *ASp* 15-18, mis en ligne le 08 mai 2012, consulté le 22 janvier 2024. URL : <http://asp.revues.org/2989> (consulté le 04 décembre 2023).

GROSSMANN Francis et TUTIN Agnès (2003). « Quelques pistes pour le traitement des collocations », in *Les Collocations : analyse et traitement. Travaux et recherches en linguistique appliquée*, Éd. De Werelt, pp. 5 -22.

JOUSSE Anne-Laure (2010), *Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales*, Thèse de doctorat, Université Paris 7-Denis Diderot.

L'HOMME Marie-Claude (2003), « Les combinaisons lexicales spécialisées (CLS) : Description lexicographique et intégration aux banques de terminologie », in *Les Collocations : analyse et traitement*, Éd. De Werelt, pp. 89-104.

MEJRI Salah (2008). « Constructions à verbes supports, collocations et les locutions verbales », in *Las construcciones verbo-nominales libres y fijas. Aproximación contrastiva y traductológica* (Dir. Mogorron Huerta, P.; Meiri, S.), pp. 191 - 202

MEL'CUK Igor (2003), « Les collocations : définition, rôle et utilité », in *Les Collocations : analyse et traitement*, Éd. De Werelt, pp. 23-32.

MEL'CUK Igor, Clas André et Polguere Alain (1995), *Introduction à la lexicologie explicative et combinatoire*, Éd. Duculot, Louvain-la-Neuve (Belgique).

POLGUERE Alain (2002), *Notions de base en lexicologie*, Observatoire de Linguistique Sens-Texte, Département de Linguistique et traduction, Université de Montréal, Canada.

POLGUERE Alain et MEL'CUK Igor (2006). « Dérivations sémantiques et collocations dans le DiCo / LAF », in *Collocations, corpus, dictionnaires. Langue Française N°150*, Éd. Larousse/Armand Colin, pp. 66-83.

PRIESOLOVA, Janka, 2018, « L'enseignement/apprentissage du lexique français des affaires dans son environnement discursif », *Synergies Europe n° 13*, pp. 127-138. URL : <https://gerflint.fr/Base/Europe13/priesolova.pdf> (consulté le 10 avril 2024).

SIEPMANN Dick (2006), « Collocations et dictionnaires d'apprentissage onomasiologiques bilingues : questions aux théoriciens et pistes pour l'avenir », in *Langue Française N° 150*, pp. 99 - 117. URL : https://www.persee.fr/doc/lfr_0023-8368_2006_num_150_2_6856 (consulté le 16 novembre 2023).

TUTIN Agnès (2013), « Les collocations lexicales : une relation essentiellement binaire définie par la relation prédicat-argument », in *Vers une extension de la phraséologie*, pp. 47 - 63.

TUTIN Agnès et GROSSMANN Francis (2002), « Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif », *Revue Française de Linguistique Appliquée*, « Lexique : recherches actuelles », Vol. VII-1, pp. 7-25. URL : <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2002-1-page-7.htm> (consulté le 12 août 2023).

VERLINDE Serge, BINON Jean, et SELVA Thierry (2006), « Corpus, collocations et dictionnaires d'apprentissage », in *Langue Française N°150*, pp. 84*98. URL : https://www.persee.fr/doc/lfr_0023-8368_2006_num_150_2_6855 (consulté le 15 octobre 2023).

WILLIAMS Geoffrey (2003), « Les collocations et l'école contextualiste britannique », in *Les Collocations : analyse et traitement*, Éd. De Werelt, pp. 33-44.

Biographie des auteurs

Aimeline RASOANANTENAINA et Cécile Marie-Ange MANOROHANTA, enseignante-chercheuses en Sciences du langage dont les travaux s'intéressent à la description linguistique des combinatoires lexicales et leurs enjeux applicatifs.