



**CE QUE LA LINGUISTIQUE DE CORPUS PEUT NOUS DIRE
SUR LES LOCUTIONS EN LANGUE FRANÇAISE ÉCRITE**
Le cas de la locution adverbiale à *première vue*.

COSIMO DE GIOVANNI
Università degli Studi di Cagliari
cosimodegiovanni@gmail.com

Abstract

(FR) Cette étude se propose d’illustrer les résultats d’une analyse menée sur un corpus de langue française, sur la locution adverbiale à *première vue*. D’un point de vue méthodologique, notre étude se situe dans l’approche de la linguistique de corpus, en nous inspirant notamment du modèle des *unités étendues de sens* ou UES (Sinclair 1991), fondée sur quatre processus d’ordre sémantique, à savoir la collocation, la colligation, la préférence sémantique et la prosodie sémantique. Dans cette optique, nous analyserons un corpus monolingue français de troisième génération afin de détecter les unités dans lesquelles la locution à première vue joue le rôle de noyau lexical. Cette étude vise à démontrer que les locutions s’inscrivent dans le cadre d’une phraséologie étendue, où les unités identifiées deviennent un principe d’organisation du texte.

MOTS-CLÉS : locutions; linguistique de corpus; corpus; phraséologie étendue; unités étendues de sens.

(EN) This study aims to illustrate the results of an analysis conducted on a French corpus, focusing on the adverbial phrase à *première vue*. From a methodological point of view, our study is based on the corpus linguistics approach, drawing inspiration, in particular, from the model of *extended units of meaning* (Sinclair 1991), based on four semantic phenomena, namely collocation, colligation, semantic preference and semantic prosody. To achieve this, we will analyze a third-generation monolingual French corpus, identifying units in which the adverbial phrase serves as the lexical core. This study aims to demonstrate that phrases are part of an extended phraseology, where the identified units serve as a principle of text organization.

KEYWORDS: locutions; corpus linguistics; corpus; extended phraseology; extended units of meaning.

1. Introduction

Désormais, il est indéniable que l’avènement de la linguistique de corpus a entraîné un changement radical dans les études linguistiques contemporaines, surtout dans le domaine de la phraséologie (cf. Williams 2006 ; Teubert 2009). En fait, les premières méthodes d’extraction automatique d’unités polylexicales à partir de grands corpus remontent au début des années 1970 (*Frantext* pour la France

et le *COBUILD* pour la Grande Bretagne)¹. Dans ces mêmes années, des approches, à la fois quantitatives et qualitatives, ont été développées pour l'étude des unités phraséologiques par excellence, les collocations.

Dans cette étude, nous nous intéressons à un autre type d'unité phraséologique, à savoir la *locution*. Issu de la grammaire traditionnelle, le terme locution reste assez controversé. En effet, dans le domaine des études linguistiques, le concept garde à la fois une acception large et une acception restreinte. Dans son acception large, le terme se superpose à d'autres termes désignant diverses unités phraséologiques, formellement et sémantiquement hétérogènes : *collocation*, *composé*, *construction à verbe support*, *expression*, *idiome*, *lexie complexe*, *phrase figée*, *phrasème*, *syntagme*, *unité polylexématique*...². Dans son acception restreinte, le terme locution, suivi d'un adjectif (nominale, verbale, adjectivale, adverbiale, prépositive, déterminative, pronominale et conjonctive), désigne un groupe de mots séparés par des blancs dans l'écriture fonctionnant comme un mot unique.

C'est la deuxième acception du terme que nous prendrons en considération dans cet article, dont l'objectif est d'examiner l'apport de la linguistique de corpus à la description des propriétés linguistiques des locutions, considérées par les dictionnaires et les grammaires comme des unités préfabriquées de la langue, avec leur propre signification (non compositionnelle) et peu productives sur le plan textuel. Pour ce faire, nous présenterons les résultats d'une analyse menée sur corpus de la locution adverbiale *à première vue*.

2. Contexte théorique et méthodologique

D'un point de vue méthodologique, notre étude se situe dans l'approche de la linguistique de corpus, en nous inspirant notamment du modèle des *unités étendues de sens* ou UES, de l'anglais *extended units of meaning*, introduit par Sinclair (1996). Ce modèle – qui est une application du principe d'idiome³ (de l'anglais *idiom principle*), développé par Sinclair (1991), – considère que le sens d'un mot dépend par des emplois co-textuels dans lesquels il apparaît⁴. Une UES n'est pas une expression figée, mais une unité lexicale abstraite régie par des contraintes d'ordre lexical, syntaxique, sémantique et pragmatique. Elle se compose d'un noyau (dénommé *nœud*) et des mots faisant partie de son environnement lexical, les *collocats*. Le nœud est lié à ses collocats au moyen de quatre processus, à savoir de la collocation, de la colligation, de la préférence sémantique et de la prosodie sémantique⁵. Les quatre processus se différencient par le type de co-sélection s'établissant entre le nœud et les autres éléments de l'environnement lexical. La collocation est la co-sélection lexicale entre le nœud et ses collocats lexicaux. La colligation est la relation régulière, préférentielle, entre le nœud et une catégorie grammaticale. La préférence sémantique est une relation de co-occurrence entre le nœud et un ensemble de mots sémantiquement liés. La prosodie sémantique est le sens connotatif, positif, négatif ou neutre, que le nœud reçoit du champ sémantique de ses collocats habituels⁶. Pour qu'une co-sélection identifiée soit intégrée dans une description, et donc considérée comme significative, elle doit être récurrente (cf. Tognini-Bonelli 1996 ; Sinclair 1999 ; Rundell 2018).

¹ À ce propos, nous renvoyons à Bernard (2002) et à Léon (2008).

² Ces termes ont été extraits des deux recueils portant sur la locution : Fiala *et al.* (1997) et Martins-Baltar (1997).

³ À ce principe s'oppose le *principe du choix libre* (de l'anglais *open choice principle*), selon lequel le locuteur dispose d'un grand choix de possibilités lexicales régi par des contraintes grammaticales (Sinclair 1991).

⁴ Cette vision s'inspire du contextualisme de Firth (1957).

⁵ Morley et Partington (2009) préfèrent parler de *signification évaluative* (de l'anglais *evaluative meaning*), tandis que Stubbs (2002) remplace le terme par celui de *prosodie du discours* (de l'anglais *discourse prosody*).

⁶ Louw définit la prosodie sémantique comme « a consistent aura of meaning with which a form is imbued by its collocates » (1993 : 157) : dans des contextes habituels négatifs, positifs ou neutres, le mot aura à son tour une prosodie sémantique négative, positive ou mixte.

Pour notre analyse, la recherche a été effectuée sur un corpus de troisième génération (ou Web-corpus⁷) dans Sketch Engine⁸ – le French Web 2023. Le corpus se compose, pour un total de presque 24 milliards de mots, de textes récupérés automatiquement sur Internet – téléchargés en décembre 2022 et janvier-février 2023 (les textes de Wikipédia français, en revanche, ont été téléchargés entre septembre et novembre 2020) –, couvrant la plus grande variété possible de genres, de sujets, de types de textes et de sources Web. Pour l’analyse des collocs, nous avons utilisé la fonction *Collocations* de Sketch Engine à partir des concordances générées – nous avons répertorié 89 341 occurrences de la locution. Une fois les paramètres sélectionnés, les collocs présents dans l’environnement du nœud (dans une fenêtre -3;3) sont classés, par défaut, selon le *logDice*, un coefficient statistique qui mesure la force d’attraction du nœud à l’égard de ses collocs (dans le texte, la mesure logDice est indiquée entre parenthèses, à côté de chaque mot)⁹.

3. Analyse des données

Pour notre étude, nous nous sommes limités à l’analyse des premiers collocs les plus fréquents selon l’indice LogDice, à savoir *sembler* (8,08), *paraître* (6,74), *paraît* (6,29), *paraissent* (5,95), *semblent* (5,64), *paraître* (5,58) et *semble* (5,39). Dans ce qui suit, nous procéderons d’abord à une analyse lexicale, puis nous essaierons d’interpréter les résultats obtenus, à travers une analyse sémantique.

3.1 *Ssembler, semblent, semble*

Avec le collocat *sembler* notre locution apparaît dans deux structures récurrentes :

- a) |pouvoir + SEMBLER + ADJ. + à première vue|
- b) |pouvoir + SEMBLER + à première vue + ADJ.|

D’après un premier examen, les deux structures varient pour la position de l’adjectif par rapport au nœud : dans le contexte de gauche, entre le nœud et le collocat dans le cas de a), dans le contexte de droite et adjacent au nœud dans le cas de b). Apparemment, il ne semble y avoir aucune raison particulière de préférer l’une ou l’autre forme. Voici quelques exemples (le soulignement est à nous) :

- (1) grâce à son système de parry, une mécanique qui peut sembler à **première vue** étrange dans
- (2) Les beercocktails peuvent sembler étranges à **première vue**, mais selon des amateurs de
- (3) Le résultat pourrait sembler à **première vue** banal, mais il n'est en réalité pas anodin.
- (4) les offrandes du chef Dyan Solomon peuvent sembler banales à **première vue** , mais tout ici

On observe également la récurrence du verbe modal *pouvoir* dans les deux structures.

Il est intéressant d’observer la récurrence de deux autres relations colligationnelles du nœud avec la conjonction de subordination, *bien que*, et la locution conjonctive *même si*. Voici quelques exemples :

- (5) Bien que cela puisse sembler relativement simple à **première vue**, le déploiement d’un réseau

⁷ Sur la notion de corpus de troisième génération, nous renvoyons à Mayaffre (2010) et De Giovanni (2021).

⁸ <https://www.sketchengine.eu/>. Pour en savoir plus sur Sketch Engine, voir Gatto (2009) et Kilgariff *et al.* (2014).

⁹ La valeur maximale de l’indice ne peut être supérieure à 14, mais elle est généralement inférieure à 10. Avec un indice de 0, la combinatoire n’est pas considérée significative (cf. Rychlý 2008).

(6) Même si un concentrateur USB peut sembler bon marché **à première vue**, il est logique de réfléchir au bon appareil

Avec le collocat *semblent*, la locution apparaît dans la structure suivante :

c) | *semblent* + ADJ./VERBE + *à première vue* |

Soit les exemples :

(7) Même les riches, qui semblent perdants **à première vue**, s'en sortiraient mieux en fin de compte

(8) , deux conceptions qui sont complémentaires mais semblent contradictoires **à première vue**.

(9) Deux affaires très récentes semblent se contredire **à première vue**. </s><s> La première,

Dans ce cas aussi, tant les adjectifs que les verbes peuvent apparaître dans le contexte de droite de la locution-nœud sans pour autant affecter le sens de l'unité :

(10) adoptent pour faire face à ces deux principes qui semblent **à première vue** contradictoires.

(11) découvrir d'une page à l'autre des réflexions qui semblent **à première vue** se contredire. </s><s>

Nous avons encore remarqué la présence, dans l'environnement lexical du nœud, de la locution conjonctive *même si* et de la conjonction de subordination *bien que*, toujours dans le contexte de gauche :

(12) Bien que les biocarburants semblent être **à première vue** une solution de remplacement pour les énergies fossiles

(13) Même si **à première vue** les caractéristiques semblent correctes, en lisant avec attention sa fiche technique

(14) Bien que ces activités réunissent des groupes qui **à première vue** ne semblent n'avoir rien en commun

(15) même si les révélations journalistiques de samedi semblent **à première vue** ne confirmer que ce qu'on savait déjà

Avec le collocat *semble*, la locution-nœud apparaît dans les mêmes structures vues précédemment. Notons cependant que, dans ce cas, la locution apparaît toujours accompagné des formes impersonnelles comme *il semble* et *cela semble* lorsqu'elle se trouve en début de phrase.

3.2 Paraître (*paraître*), *paraît*, *paraissent*

Avec le collocat *paraître* (et sa variante orthographique *paraître*), la locution est présente dans les deux structures suivantes :

d) |pouvoir + PARAÎTRE + ADJ. + *à première vue* |

e) |pouvoir + PARAÎTRE + *à première vue* + ADJ. |

Même dans ce cas, le changement de position des éléments, en particulier des adjectifs, ne semble avoir aucune implication sur la signification de la structure. Voici quelques exemples :

(16) Cette seconde option peut paraître paradoxale à **première vue**, mais elle résulte d'un certain constat d'impuissance des promus

(17) Cela peut paraître à **première vue** paradoxal, pourtant de nombreux vacanciers partent seuls en voyage organisé.

En présence du collocat *paraître*, la locution-nœud entretient des relations colligationnelles avec *même si* et *bien que*.

Avec le collocat *paraît*, notre locution est à la base des mêmes structures repérées pour le collocat précédent. Cependant, nous y découvrons une construction récurrente (634 occurrences), à savoir [qu'il n'y paraît + à *première vue*], comme dans les exemples suivants :

(18) Un thème riche, bien plus ouvert qu'il n'y paraît à **première vue** et qui laisse libre cours à votre imagination

(19) En fait, l'alphabet coréen n'est pas aussi compliqué qu'il n'y paraît à **première vue** par rapport aux autres langues asiatiques.

Nous avons rencontré la même structure avec le collocat *paraissent*, dans la forme *qu'ils n'y paraissent*, mais avec une fréquence plus modeste (7 occurrences).

4. Interprétation des résultats et conclusions partielles

L'analyse des sept premiers collocats de à *première vue*, classés par logDice, nous a permis d'esquisser un profil collocationnel et un profil colligationnel de la locution-nœud. Concernant le profil collocationnel, la locution est employée préférentiellement avec les verbes *sembler* et *paraître*. En ce qui concerne le profil colligationnel, à l'intérieur de son environnement lexical, le nœud est précédé d'un verbe modal (*pouvoir*), d'une locution conjonctive (*même si*) et d'une conjonction de subordination (*bien que*). Cependant, cette analyse seule ne suffit pas. Une analyse de la dimension sémantique et des réalisations pragmatiques des unités identifiées s'impose, à travers l'interprétation des données.

Dans les contextes linguistiques analysés, la locution à *première vue* prend le sens de quelque chose ou quelqu'un qui apparaît différemment de ce qu'il est ou pourrait être ; cette interprétation est conforme à ce qui est décrit dans les dictionnaires de langues. D'un point de vue discursif, elle exprime un jugement, voire un point de vue, provisoire de la part du locuteur. C'est le résultat d'une première impression, qu'elle soit positive ou négative (ce qui dépend de l'utilisation des adjectifs), à l'intérieur du processus énonciatif. La présence des deux modaux *sembler* et *paraître* signifie une intervention du locuteur par rapport à l'énonciation, c'est-à-dire le caractère provisoire de son jugement. C'est précisément grâce à sa co-présence avec l'un des deux verbes que la locution adverbiale montre sa préférence sémantique pour l'apparence

Sur le plan de la prosodie sémantique, nous pouvons affirmer que la locution adverbiale à *première vue* a à peu près le même nombre d'occurrences dans un environnement comportant une évaluation à la fois positive et négative. Cela signifie que le jugement initial du locuteur, ce qui a un caractère provisoire, au moment de l'énonciation, peut être caractérisé par une différente connotation, selon les cas.

Cependant, sur le plan énonciatif, le jugement ne reste pas provisoire, le point de vue du locuteur ne s'arrête pas à la première impression, mais il subit une évolution. Cette évolution est garantie par

la présence, dans le contexte de droite de la locution, de conjonctions adversatives (*mais, toutefois, cependant*) et à un changement de jugement par l'utilisation d'éléments lexicaux mis en relation avec les éléments de la première partie de l'énoncé, ce qui implique le passage à un jugement stable. Nous proposons quelques exemples toujours tirés du même corpus¹⁰ :

(20) Choisir la bonne façon d'accrocher, et monter des rideaux peut sembler assez compliqué à **première vue**, mais bien souvent cela n'est pas aussi difficile que ça en a l'air.

(22) Certaines annonces d'emplois paraissent sérieuses à **première vue**, cependant, certains éléments peuvent se révéler suspects.

(23) Ces dimensions qui semblent réduites à **première vue** sont toutefois capables d'accueillir un PC de bureau ou un portable

Des exemples de ce type sont présents dans tout les occurrences du corpus où la locution adverbiale et les conjonctions adversatives constituent ensemble une unité lexicale stable marquant un avant et un après dans le processus énonciatif.

Le même processus de changement du jugement a lieu en présence de la locution conjonctive *même si* et de la conjonction de subordination *bien que* (5)-(6) et (12)-(15). Ce même sens est contenu dans l'expression récurrente *qu'il n'y paraît à première vue* que l'on peut considérer une unité phraséologique à part entière, ou bien une séquence figée, caractérisée par une fixité formelle, sémantique et pragmatique.

Cette brève analyse nous a permis d'observer l'emploi de la locution à *première vue* dans un corpus composé de textes de divers genres, sujets, et types tirés du Web, entre 2020 et 2023. L'approche que nous avons adoptée a le mérite d'avoir mis en évidence une partie de son comportement réel dans la structure des textes. Contrairement à ce que rapportent les dictionnaires et les grammaires, en nous appuyant sur l'approche de la linguistique de corpus, nous avons pu démontrer que la locution est une unité dynamique de la langue qui co-participe à la formation du sens, au sein de structures lexicales plus étendues.

Références bibliographiques

- Bernard P., 2002, "L'ATILF et ses ressources linguistiques informatisées", *Zeitschrift Für Französische Sprache Und Literatur*, 112 : 69-79. <http://www.jstor.org/stable/40599464> (8.02.2024).
- De Giovanni C., 2021, "Corpus, Web et recherche terminologique : le cas des termes du domaine du bien-être animal", *Synergies Italie*, 17 : 67-77.
- Fiala P. et al., 1997, *La locution : entre lexique, syntaxe et pragmatique. Identification en corpus, traitement, apprentissage*, Klincksieck, Paris.
- Firth J.-R., 1957, *Papers in Linguistics (1934-1951)*, Oxford University Press, Oxford.
- Gatto M., 2009, *From Body to Web. An Introduction to the Web as Corpus*, Laterza, Bari.
- Kilgariff A. et al., 2014, "The Sketch Engine: ten years on", *Lexicography*, 1 : 7-36.
- Léon J., 2008, "Aux sources de la 'Corpus Linguistics' : Firth et la London School", *Langages*, 171 : 12-33.
- Louw W.E., 1993, "Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies", in M. Baker, G. Francis, E. Tognini-Bonelli (eds), *Text and Technology: In Honour of John Sinclair*, Benjamins, Amsterdam : 157-176.

¹⁰ Pour mieux illustrer les caractéristiques sémantico-pragmatiques des unités détectées, nous avons été obligés d'élargir les segments de texte.

- Martins-Baltar M., 1997, *La locution entre langue et usages*, ENS Éditions, Lyon.
- Mayaffre D., 2010, “Corpus et web-corpus. Réflexion sur la corporalité numérique”, *Cahiers de praxématique*, 54-55 : 233-248.
- Morley J., Partington A., 2009, “A few Frequently Asked Questions about semantic – or evaluative – prosody”, *International Journal of Corpus Linguistics*, 14 : 139-158.
- Rundell M., 2018, “Searching for extended units of meaning – and what to do when you find them”, *Lexicography ASIALEX*, 5 : 5-21. DOI : <https://doi.org/10.1007/s40607-018-0042-1> (8.02.2024).
- Rychlý P., 2008, “A Lexicographer-Friendly Association Score”, in P. Sojka, A. Horák (eds), *Proceedings of Recent Advances in Slavonic Natural Language Processing*, Masaryk University, Brno : 6-9.
- Sinclair J., 1991, *Corpus, Concordance, Collocation*, Oxford University Press, Oxford.
- Sinclair J., 1996, “The Search for Units of Meaning”, *Textus*, 9 : 75-106.
- Sinclair J., 1999, *The computer, the corpus and the theory of language* : <https://www.openstarts.units.it/server/api/core/bitstreams/1fa0ecd1-90c1-4695-99fa-4335782b54e9/content> (8.02.2024).
- Stubbs M., 2002, *Words and Phrases. Corpus Studies of Lexical Semantics*, Blackwell, Oxford.
- Teubert W., 2009, “La linguistique de corpus : une alternative”, *Semen. Revue de sémio-linguistique des textes et discours*, 27. DOI : <https://doi.org/10.4000/semen.8923> (8.02.2024).
- Tognini-Bonelli E., 1996, “Translation Equivalence in a Corpus Linguistics Framework”, *Rivista internazionale di tecnica della traduzione*, 2 : 28-53.
- Tognini-Bonelli E., 2001, *Corpus Linguistics at Work*, John Benjamins Publishing Company, Amsterdam.
- Williams G., 2006, *La linguistique de corpus : une affaire prépositionnelle*, in C. Duteil-Mougel, B. Foulquié (eds), *Corpus en Lettres et Sciences sociales. Des documents numériques à l'interprétation. Actes du Colloque international et école d'été Albi, 10-14 juillet 2006 Organisé dans le cadre des Colloques d'Albi Langages et Signification (CALS)* : http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Actes_ALBI-06.pdf (8.02.2024).