# Un corpus pour l'enseignement des collocations en FLE

Rui Yan, Thi Thu Hoai Tran, Cristelle Cavalla

▶ **To cite this version:**

Rui Yan, Thi Thu Hoai Tran, Cristelle Cavalla. Un corpus pour l'enseignement des collocations en FLE. Etudes en didactique des langues, 2018, Les corpus / Corpora, 31, pp.93-107. hal-04034813

**HAL Id: hal-04034813**

**https://ut3-toulouseinp.hal.science/hal-04034813**

Submitted on 17 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Un corpus pour l'enseignement des collocations en FLE

*Rui YAN*, lectrice, Lidilem, Université Grenoble Alpes
*Thi Thu Hoai TRAN*, MCF, Grammatica, Université d'Artois
*Cristelle CAVALLA,* professeure, Diltec, Université Sorbonne-Nouvelle Paris 3

## Introduction

Notre travail se situe autour d'une question maintes fois débattue: comment peut-on faire entrer les corpus en classe de langue (Boulton & Tyne, 2014; Chambers, 2007; Fligelstone, 1993; Johns, 2002)? Par les collocations, nous montrerons plusieurs tentatives didactiques d'introduction des corpus en classe de FLE. Nous présenterons des exemples d'activités développées à l'aide du corpus Scientext et de son application "ScienQuest" pour l'aide à l'écrit scientifique auprès d'étudiants allophones dans les universités françaises.

Nous nous intéressons tout particulièrement ici aux collocations. Il s'agit en effet d'expressions binaires de type: *jouer un rôle, résultat convaincant*. Ce phénomène phraséologique, qui constitue des motifs lexico-grammaticaux propres aux discours spécialisés, a fait l'objet de nombreuses études (Gledhill, 2011; Grossmann & Tutin, 2003). En fait, les collocations sont omniprésentes dans le discours des natifs mais constituent une source de difficulté pour les étudiants allophones (Nesselhauf, 2005; Bolly, 2011; Cavalla, 2015). Dans une étude récente (Yan & Tran, 2018), nous avons remarqué que, d'une part, les collocations erronées sont bien présentes dans les écrits des étudiants et que, d'autre part, les collocations spécifiques à l'écrit scientifique sont peu présentes chez les étudiants allophones tandis qu'elles sont fréquentes chez les natifs experts du domaine. Cette différence apparaît pour des collocations telles que *soulever un problème*, *s'appuyer sur une théorie* ou encore *conduire à une conclusion*. Il nous semble donc important de développer des enseignements autour des collocations.

Dans cette recherche, nous envisageons d'enseigner des collocations scientifiques transdisciplinaires et, pour ce faire, d'introduire des corpus numériques en classe. L'introduction du corpus se fera dans le cadre d'une aide à l'écrit universitaire pour des étudiants allophones. La question qui pourrait résumer notre recherche actuelle est la suivante: comment enseigner les collocations scientifiques à l'aide de corpus en classe de FLE?

Toutefois, avant de présenter quelques réponses, nous essaierons de définir ce qu'est une collocation pour l'enseignement en français universitaire auprès d'allophones et nous verrons les liens entre corpus et collocations. Car, pourquoi enseigner ces éléments linguistiques avec un outil comme les corpus au vu de l'interface de certains, encore très rébarbative?

**Qu'est-ce qu'une collocation?**

Les collocations sont des éléments fréquents dans la langue, comme l'atteste Sinclair (1991: 112) à propos de l'anglais mais que nous vérifions pour le français dans les corpus (Scientext[1] entre autres).

> Les collocations sont apparues partout dans les concordances générées à partir du corpus. L'idiomaticité, loin d'être un aspect marginal de la langue, semble être omniprésente et au moins aussi importante que la syntaxe dans la construction du sens. (Traduction des auteures[2].)

De fait, il apparaît important d'enseigner ces éléments à des étudiants allophones désireux de suivre des cursus universitaires dans des universités françaises. En outre, l'université française, telle qu'elle fonctionne aujourd'hui, demande de nombreux écrits dans lesquels ces structures langagières sont très présentes, tant dans le lexique spécifique à chaque discipline que dans le lexique transversal à toutes les sciences enseignées, à savoir le lexique transdisciplinaire des écrits scientifiques (Tutin, 2007). Ajoutons que, au vu des recherches en psycholinguistique, l'enseignement de ces éléments langagiers devient indispensable dès l'instant où l'on admet que:

> Un grand nombre d'expériences a essayé de démontrer qu'une séquence dite formulaire (et, donc, stockée et extraite en bloc) est en effet appréhendée plus rapidement qu'une séquence similaire, mais non phraséologique (Edmonds, 2013: 125).

Edmonds (2013) rappelle que la fréquence est à prendre en compte à la fois pour le stockage et la rapidité de production pour les allophones. Outre ces considérations permettant de prouver l'utilité de l'enseignement de ces éléments en FLE, ce qui importe alors est la définition de ces éléments que nous pouvons retenir pour l'enseignement. Quand nous évoquons "l'enseignement", il s'agit d'aiguiller l'apprenant dans la jungle des éléments linguistiques à disposition dans la langue française. De fait, et parce que peu d'étudiants face à nous sont linguistes, il convient de trouver une description simple d'accès pour des non spécialistes afin qu'ils puissent reconnaître aisément les structures en question.

La communauté des linguistes considère généralement, la collocation comme une association fréquente de deux unités lexicales, en d'autres termes, une

[1] ANR Scientext, laboratoire Lidilem. Responsables: Agnès Tutin & Francis Grossmann. URL: http://scientext.msh-alpes.fr/scientext-site/spip.php?article1.

[2] Collocational patterns appeared everywhere in the lexical concordances generated from the corpus. Idiomaticity, far from being a marginal aspect of language, seemed to be ubiquitous and at least as significant as syntax in the construction of meaning (Sinclair, 1991: 112).

cooccurrence de deux constituants linguistiques entretenant une relation sémantique et syntaxique privilégiée (Cavalla & Sorba, 2017; Grossmann & Tutin, 2003; Hausmann 1989; Larivière 1998; Polguère, 2008; Tutin, 2013). Dans cette association, les deux lexies ont un statut inégal: une base autonome au plan sémantique est unie à un collocatif qui lui ajoute une caractérisation sans en modifier l'identité et qui voit son sens usuel ainsi modifié. Par exemple, dans *semer le doute* ou *poser une question*, les verbes *semer* et *poser* sont des collocatifs, c'est-à-dire des éléments dont le sens usuel est modifié tandis que *doute* et *question* sont les bases dont le sens usuel reste le même.

Des définitions plus statistiques, comme celles de Firth (1957) et Halliday et *al.* (1964), prennent en compte des critères de fréquence, la fréquence restant, pour ces contextualistes, le critère définitoire essentiel (Sinclair et *al.*, 2004). Cette optique est suivie, par exemple, par Williams (2006), Tognini-Bonelli (2001) pour l'aide à la rédaction universitaire, ou encore Gavioli (2005).

Ainsi, avons-nous deux entrées définitoires qui – de notre point de vue d'enseignantes de FLE – se complètent pour l'enseignement/apprentissage du français universitaire: d'un côté, une vision sémantique avec des éléments dont le sens peut être plus ou moins affecté en fonction de leur association; et, de l'autre, une haute fréquence d'apparition des éléments dans des contextes précis[3]. Dans les deux cas, ces combinaisons lexicales peuvent admettre des insertions lexicales, voire des substitutions plus ou moins synonymiques. Prenons deux exemples.

- *Principales caractéristiques* apparaît dans des textes d'économie dans le corpus Scientext; sa fréquence est suffisamment élevée pour en tenir compte dans l'enseignement. Toutefois, l'aspect sémantique d'une base et surtout d'un collocatif qui prendrait un sens non-usuel n'apparaît pas dans cette association. De fait, il est préférable ici de s'attacher à la définition statistique.
- *Piste de recherche* apparaît dans différentes disciplines des sciences humaines (linguistique, sciences de l'éducation, psychologie) dans Scientext. Cette collocation est rare dans ce corpus (22 occurrences), mais suffisamment pertinente en sciences humaines pour qu'on y prête attention dans l'enseignement. En outre, nous pouvons clairement identifier une base (*recherche*) et un collocatif (*piste*) dont le sens usuel est quelque peu modifié[4].

Finalement, que retenir pour l'enseignement? Nous rencontrons des collocations répondant aux deux définitions dans les écrits des apprenants. Les collocations utilisées de façon erronée nous obligent à nous arrêter sur ces éléments. Ne pouvant donner des définitions trop spécialisées à des non spécialistes, nous optons pour conserver les deux entrées, sans les nommer, mais en leur présentant les caractéristiques saillantes des constructions, notamment au plan

[3] En effet, les genres influencent le choix d'utilisation de certaines collocations.

[4] Dans le TLFi (*Trésor de la langue française informatisé*), la "piste" désigne une "suite d'empreintes laissées par un être vivant sur le sol où il s'est déplacé", si l'on s'accorde toutefois à envisager que cette définition est plus courante que "voie suivie ou à suivre pour mener à bien une recherche, faire une découverte" qui est considéré comme un dérivé métonymique de l'acception précédente.

sémantique. En effet, comme le rappelle Médioni (2009: 20), "Un traitement qui fait intervenir le sens s'opère à un niveau plus profond: par exemple, des tâches plus difficiles qui font appel à des facultés de raisonnement ou à la comparaison détaillée".

Voici quelques exemples rencontrés dans des écrits d'étudiants allophones.

- "Si un élève ne participe pas, il ne peut pas améliorer ou *grandir dans sa langue seconde*" (étudiant anglophone, niveau B2): définition sémantique pour *évoluer/avancer dans sa langue*.
- "Ils faisaient *un vacarme de bruit* pour m'empêcher de […]" (étudiant sinophone, niveau B2): la fréquence statistique de l'apparition de l'association *vacarme assourdissant* permet de refuser *vacarme de bruit*.

Au vu de ces éléments, nous retiendrons d'abord la fréquence en réception et production, qui va de pair avec la reformulation (Hilton, 2005; Van der Linden, 2006). Même si les didacticiens des langues constatent au quotidien que la fréquence d'utilisation (réception et production pour l'apprenant) d'un élément aide à sa mémorisation, nous n'avons pas trouvé d'étude permettant de montrer comment se développe la mémorisation via la fréquence et la reformulation. En d'autres termes, existe-t-il une corrélation entre le type de reformulation et la fréquence de réception/production? Parmi les stratégies d'apprentissage, quelle serait la reformulation la plus efficace pour aider l'apprenant de L2 à retenir les collocations?

Nous ne pouvons nier le fait que plus nous entendons et/ou lisons un mot et mieux nous le retenons. Ajoutons que le critère du contexte, avancé par Halliday *et al.* (1964), favorise l'apprentissage du lexique (Ellis, 1997). Ces deux critères combinés permettent de comprendre qu'un apprenant retiendra mieux *poser une question* que les termes *poser* ou *question* isolés, car il entendra plus souvent – dans un contexte d'enseignement – la collocation et donc l'association des deux éléments – leur co-occurrence immédiate – dans une situation de communication spécifique donnée: l'université.

Pour l'enseignement des collocations dans le domaine du FLE, nous retiendrons les deux tendances présentées ci-dessus: d'un côté, celle qui s'appuie sur une description sémantique et, de l'autre, celle qui prône une description lexicale. Les lexicographes comme Hausmann (1985) et Mel'cuk *et al.* (1995) proposent d'analyser ces unités lexicales en partant de la relation sémantique entre les deux éléments dont la "base" garde son sens habituel, et le "collocatif" a un sens déviant; par exemple, *prendre un risque* est particulièrement utilisé en finance et dans la presse dans le corpus Emolex[5]. Par ailleurs, d'autres linguistes se fondent sur des critères statistiques pour identifier les collocations. Il s'agit des groupes de mots particulièrement fréquents dans les écrits, comme *résultats similaires* qui est représentatif dans les thèses en biologie (voir corpus Scientext).

Les corpus numériques qui aident à constater ces usages seront donc privilégiés ici dans un cadre d'enseignement universitaire pour l'aide à la rédaction

[5] http://phraseotext.u-grenoble3.fr/emoBase/.

académique. L'utilisation des corpus représente un réel avantage pour l'enseignant et les étudiants allophones: d'une part, l'enseignant peut sélectionner les collocations les plus saillantes et les plus fréquentes dans les écrits scientifiques; d'autre part, les étudiants peuvent travailler sur des données authentiques et apprendre à se familiariser avec les normes rédactionnelles d'un genre spécifique.

## Pour l'enseignement/apprentissage des collocations à l'aide de corpus

*Contexte et public*

Les séquences proposées s'adressent aux étudiants allophones de deuxième ou troisième cycle dans les universités françaises ou francophones qui ont à rédiger un écrit scientifique (rapport de recherche, mémoire, thèse ou article de recherche). Ces étudiants, venant de toutes les disciplines scientifiques, ont un niveau avancé en français (B1/B2), mais rencontrent diverses difficultés rédactionnelles, en particulier celles qui sont liées à la phraséologie spécifique et, plus précisément, aux collocations (Cavalla, 2015). Ces éléments lexicaux se révèlent être l'une des principales sources des nombreuses erreurs chez les apprenants de FLE.

*Objectifs et corpus*

Rappelons que le lexique transdisciplinaire est fréquent dans les écrits académiques, qu'il a un rôle dans la cohérence discursive, qu'il intègre une importante phraséologie et qu'il assure des fonctions métalinguistiques et métadiscursives (Tutin & Grossmann, 2014), communes à toutes les disciplines. Afin de mettre en évidence les collocations transdisciplinaires, nous nous basons sur le corpus Scientext[6] qui est cohérent avec nos objectifs linguistiques. Ce corpus spécialisé de cinq millions de mots regroupe des écrits scientifiques venant de dix disciplines de sciences humaines et sociales telles que: la linguistique, la psychologie, les sciences de l'éducation, l'économie, les sciences politiques, l'anthropologie, l'histoire, la géographie, les sciences de l'information et de la communication et la sociologie. Il a été balisé en parties textuelles (introduction, conclusion, notes de bas de page, etc.), et mis au format XML-TEI[7] (Falaise, 2014). Par ailleurs, il a été annoté aux niveaux morphologique et syntaxique à l'aide de l'analyseur en dépendances XIP[8] (Aït-Mokhtar *et al.*, 2002). Ces annotations permettent d'extraire les collocations. Le Tableau 1 ci-dessous illustre le repérage des collocations [*poser* + N] dans le corpus Scientext grâce aux concordanciers.

[6] Ce corpus est accessible en ligne: https://corpora.aiakide.net/scientext19/.

[7] TEI: Text Encoding Initiative désigne un ensemble de normes pour l'encodage des textes.

[8] XIP: Xeros Incremental Parser a été développé à XRCE Xeros Research Center Europe à Grenoble et permet d'effectuer une analyse morpho-syntaxique.

| Contexte gauche | Occurrence | Contexte droit |
|---|---|---|
| Les actions des évêques ruthènes de 1595-1596 n'avaient pas remis en cause les moyens et les voies du salut, mais elles | *posèrent clairement la question* | de l'autorité responsable de conférer le sacrement de l'ordre. [histoire_18636] |
| Les circonstants de prédicat sont rattachés au verbe, ils font partie du prédicat qu'ils spécifient sous un ou plusieurs rapports. Ces spécifications, plus ou moins prévisibles et attendues selon le verbe peuvent | *poser des problèmes* | de délimitation avec les compléments essentiels. [linguistique_110] |

*Tableau 1* **– Concordancier des collocations [*poser* + N]**

Dans le corpus Scientext, nous avons relevé deux collocations: *poser* + *question* (185 occurrences) et *poser* + *problème* (141 occurrences). Les bases *question* et *problème* gardent leur sens alors que le sens du collocatif *poser* est modifié. En outre, ces deux collocations sont de haute fréquence, ce qui donne à l'enseignant toutes les raisons pour les conserver: ces collocations sont repérables par le sémantisme quelque peu déviant de l'un de leurs éléments (ici le verbe *poser*) et elles sont très fréquentes dans ce type de discours. Nous pouvons ajouter qu'elles acceptent quelques insertions lexicales comme *poser clairement la question*, mais peu de substitutions lexicales sont possibles (*poser un souci* pourrait se substituer à *poser une question* dans certains contextes).

*Approche didactique inscrite dans la linguistique de corpus*

Les développements de la linguistique de corpus ont contribué au renouvellement de la réflexion sur les contenus d'enseignement et sur la place des connaissances linguistiques dans l'enseignement (Pecman, 2005). Grâce aux corpus, on peut observer les phénomènes phraséologiques de la langue, ce qui présente un intérêt certain pour l'enseignement de la phraséologie. Notre approche didactique s'inscrit dans les travaux de recherche en *data-driven learning* (DDL) initiés par Johns (1991). Le DDL vise à donner aux apprenants un accès direct aux données linguistiques et à les placer au cœur de l'apprentissage, ce qui leur permet d'endosser le rôle du "détective" et de construire, par conséquent, leurs propres savoirs de façon inductive. En fait, l'utilisation des corpus pour la didactique est une application importante et ancienne. De nombreux chercheurs s'accordent sur les avantages d'introduire les corpus en classe (Chambers, 2005; Boulton, 2008). L'enseignant peut trier des exemples authentiques et variés pour constituer les matériaux didactiques. Il peut s'agir d'exemples sur papier ou en ligne. L'apprenant, quant à lui, avec l'utilisation des corpus, pourra développer l'autonomie de la recherche en ligne, car il peut rapidement trouver seul les éléments linguistiques qui le préoccupent personnellement.

Néanmoins, il nous semble important de souligner que l'utilisation effective des corpus en classe constitue encore un défi, notamment pour les enseignants. En

effet, elle nécessite des formations pour la manipulation des corpus numériques ainsi qu'un changement du rôle de l'enseignant face aux apprenants. Le rôle de guide de l'enseignant prend alors tout son sens dès l'instant où l'autonomie de l'apprenant est visée. En ce sens, le guidage de l'apprenant relève d'une activité réflexive qui débouche sur des questionnements métalinguistiques, notamment dans l'observation des données du corpus. Il faudrait donc une formation à la méthodologie scientifique de la linguistique de corpus pour les enseignants. Nous allons montrer cet aspect de réflexion métalinguistique dans les activités didactiques présentées ci-après.

Le corpus Scientext est déjà utilisé dans plusieurs études empiriques pour enseigner l'écrit scientifique (Cavalla, 2015; Tran, 2014; Yan, 2017). Les résultats de ces études sont encourageants. Inscrites dans la même lignée de ces travaux, les activités proposées ci-dessous visent à montrer l'intérêt de l'utilisation du corpus Scientext pour enseigner et apprendre les collocations spécifiques et donner des pistes pour les enseignants de FLE.

## Réflexions autour des séquences

Pour un enseignement systématique des collocations, plusieurs aspects pourront être abordés: les aspects lexico-syntaxiques du phénomène avec, par exemple, une aide au repérage des collocations; les aspects syntagmatiques qui conduisent les apprenants à se pencher sur les constructions spécifiques autour des collocations (nominalisation, insertion d'éléments spécifiques, etc.); les aspects sémantico-discursifs, c'est-à-dire, faire travailler sur le sens associé à la place discursive de certaines collocations.

### *Aspects lexico-syntaxique*

Cette étape s'avère importante pour aider les apprenants à repérer les collocations, la structure des associations lexicales spécifiques, car il est difficile d'identifier ce phénomène sans connaître sa structure. Nous proposons d'utiliser des exercices de repérage pour travailler cet aspect des collocations. Ce type d'activité permet aux apprenants de prendre conscience de l'association quasi-indéfectible des deux éléments, par exemple: *prendre* + *conscience* ou *jouer* + déterminant + *rôle*. L'enseignant peut poser différentes questions pour aider les apprenants à repérer les éléments qui se trouvent avant et après les noms.

> **Consigne**: *Observez les expressions binaires qui servent à exprimer une posture d'auteur. Observez leur structure syntaxique. Que dire de la place des éléments entre eux? Peut-on modifier cette place? Émettez des hypothèses.*
>
> 1. Ce critère permet *d'abandonner une hypothèse* en cours si la mesure de qualité descend en dessous de celles obtenues pour les hypothèses déjà traitées (cas de segment caché par exemple).
>
> 2. Il est toujours possible d'ajouter des hypothèses *ad hoc*, pour sauver une loi. Cela signifie qu'à partir du moment où l'on souhaite *rejeter une théorie* scientifique, on doit faire intervenir un jugement de valeur.

> 3. Mais ce réarrangement cognitif est très lourd car le cerveau doit accepter *d'abandonner une idée* ancienne et réorganiser tous les savoirs du domaine en fonction de la nouvelle représentation.

Outre les exercices qui invitent à la réflexion, l'enseignant peut proposer des exercices lacunaires afin de sensibiliser les apprenants aux différentes nuances qu'apportent les adverbes. Les exemples qui suivent portent sur les collocations de démarcation. L'utilisation de différents adverbes permet de relativiser la volonté de démarcation de l'auteur. Les adverbes tels que *quelque peu*, *un peu*, accompagnent la démarcation partielle ou nuancée, tandis que l'adverbe *totalement* complète la démarcation totale.

> ***Consigne**: Complétez les phrases ci-dessous par l'adverbe TOTALEMENT, QUELQUE PEU ou UN PEU, pour mieux illustrer la volonté de se démarquer de l'auteur.*
>
> 1. Loin de contester [*totalement*] un principe d'internationalisation du travail, les tenants de la bouteille à moitié pleine considèrent que le top management confond vitesse et précipitation, prenant ses désirs de mondialisation pure et parfaite pour la réalité.
>
> Les fonctions inverses du schéma ne sont pas mobilisées explicitement ici: elles viennent généralement renforcer l'absence de fonction positive, mais elles sont aussi importantes pour relativiser [*quelque peu*] une approche positiviste. Ces fonctions négatives sont toujours plus ou moins présentes.

Rappelons que cette première phase a pour objectif la prise de conscience de l'existence de ces associations lexicales et donc l'aide à leur repérage dans les phrases dans un premier temps. Nous avançons ensuite vers le sens.

*Aspects syntagmatiques*

L'enseignement des aspects syntagmatiques a pour objectif de guider l'apprenant en partant de l'observation des collocations pour aller vers les comportements lexico-syntaxiques du mot, en particulier du point de vue transformatif. Il s'agit donc de s'approprier le mot dans un contexte plus élargi.

Il est souhaitable de commencer par des associations lexicales ou des constructions que les apprenants connaissent et de passer ensuite au repérage des modifications ou transformations aux niveaux lexico-syntaxiques. Il est important de guider les apprenants vers une comparaison pour qu'ils comprennent les effets produits au plan sémantique notamment.

Plusieurs types de modifications lexico-syntaxiques sont envisageables.

- *Modification par insertion adverbiale*: "L'interaction de ces phénomènes pose *pleinement* des problèmes auxquels sont confrontés [...]".
- *Modification par nominalisation*: "*L'interprétation* des résultats sera présentée en deux parties [...]".
- *Modification par passivation*: "La *section suivante est consacrée* à la présentation des résultats [...]".

Pendant cette étape de repérage, l'enseignant peut amener les apprenants à découvrir que le mot a une préférence parmi les différentes constructions possibles, par exemple le verbe *consacrer* s'associe fréquemment avec le passif.

Après le repérage, l'enseignant peut proposer aux apprenants de noter les associations lexico-syntaxiques rencontrées autour d'un sens déjà étudié et de compléter la liste avec d'autres constructions à l'aide du corpus afin de fixer leurs connaissances. Par exemple, pour le sens "validation d'une hypothèse", on trouve également *confirmer partiellement notre hypothèse*, *vérifier certaines hypothèses*, *l'hypothèse est validée*. Rappelons que les constructions doivent être acquises aux trois niveaux: sémantique, lexical et syntaxique.

L'étape suivante consiste à vérifier si les apprenants sont capables de manipuler ces variations lexico-syntaxiques dans un autre contexte. Il est possible de réfléchir à un ensemble de synonymes pour vérifier, non seulement la bonne mémorisation des constructions, mais aussi la bonne compréhension du sens des éléments. En effet, certaines bases n'acceptent qu'un nombre limité de substitutions. Nous proposons ci-dessous un exemple d'exercice portant sur les collocations V + *exemple*.

***Consigne**: Repérez le sujet des verbes "citer" et "donner" et comparez les deux phrases (à gauche et à droite). Quelle est la différence au niveau du verbe? Quel est l'effet produit?*

| **Phrase** | **Phrase à comparer** |
|---|---|
| *Citons* juste *un exemple*, que nous avons observé sur les deux pupitres étudiés ici. | Damourette et Pichon en *citent quelques exemples* qu'ils auraient entendus oralement, et qui paraissent assez isolés [...] |
| Le seul *exemple* qui établisse un lien de causalité *est donné* par le NPR: il a reçu des reproches au sujet de sa conduite où la cause des reproches est bien due à la conduite de "il". | Nous en *donnons quelques exemples* qui seront commentés de manière succincte. |
| Cette approche combinant différents types de standards a contribué, au moment de la transposition en droit national et de la mise en œuvre, au renforcement des approches nationales en matière de protection de l'environnement, *comme le montre l'exemple* de la lutte contre la pollution. | *Cet exemple montre* qu'on peut avoir besoin de conserver la notion de sens littéral, auquel le locuteur réduit souvent sa responsabilité en cas d'implicitation contestée [...] |

Cet exercice invite les apprenants à comprendre, non seulement le sens du verbe, mais aussi à repérer le type de construction dans laquelle il apparaît. L'enseignant peut guider les apprenants à rappeler les types de constructions par exemple: *le seul exemple est donné* ou *nous donnons le seul exemple*. Par la suite, il sera important que les apprenants soient conscients des différences sémantiques entre ces constructions. Par exemple, la construction en *comme* dans *comme le montre l'exemple* est employée pour guider le lecteur dans le constat des faits et renforcer l'argumentation.

*Aspects sémantico-discursifs*

Afin de sensibiliser les apprenants aux aspects sémantiques des collocations, deux grandes catégories d'exercices peuvent être proposées, soit sur la synonymie soit sur la fonction rhétorique.

Les questions d'un premier exercice peuvent porter sur l'aspect sémantique de la base et du collocatif. Nous pouvons faire observer la phrase suivante: *Les résultats obtenus corroborent et précisent notre hypothèse de travail initiale.*

Les questions pour conduire au repérage sémantique portent sur le sujet et le verbe. Par exemple: Qu'est-ce que les résultats corroborent? Qu'est-ce que les résultats précisent? Qu'est-ce que l'auteur fait avec son hypothèse?

D'autres questions conduisant à une plus grande précision sémantique peuvent être envisagées.

> *Parmi les verbes suivants, lequel pourrait-on utiliser après le nom "hypothèse" comme synonyme de "préciser"? (valider, spécifier, souligner, servir)*
>
> *Pourrait-on proposer autre chose que "hypothèse" parmi les noms suivants: thèse, postulat, question, problématique?*
>
> *À votre avis, pourquoi certains verbes puis certains noms ne sont-ils pas possibles? Par exemple, pourquoi le verbe "dire" n'est-il pas possible?*
>
> *Faites la liste des verbes possibles autour de "hypothèse".*

Après cette première étape, des exercices de reformulation lexicale peuvent être proposés, ce qui sensibilise les apprenants à la fonction rhétorique des collocations au sein du discours, et leur permet de varier les expressions dans leurs écrits. Les exemples suivants portent sur les collocations de démarcation: les étudiants doivent utiliser des collocations de démarcation pour donner plus de force démarcative au propos.

> ***Consigne**: Réécrivez les phrases suivantes en utilisant des collocations de démarcation synonymes, pour donner plus de force démarcative au propos.*
>
| **Phrases** | **Reformulation proposée** |
|---|---|
| *Cette hypothèse semble aller à l'encontre d'une série d'expériences* faites par Varley et son équipe […] qui montrent que des aphasiques agrammatiques peuvent […] | Cette *hypothèse semble contredite* par une série […] |
| *Il n'y a rien de mal à ne pas suivre un modèle*, d'autant que la France et l'Europe n'ont aucunement besoin d'un modèle autre que le leur. | … Il n'y a rien de mal à *rejeter un modèle* […] |

Travailler les aspects sémantiques par l'entrée de fonctions rhétoriques impose aux apprenants de comprendre simultanément le sens des collocations en lien avec la place qu'elles peuvent avoir dans un écrit, ici scientifique. Par exemple, pour la phrase "Nous reprenons l'hypothèse de Parker (1970)", la valeur sémantique associée est de présenter, à travers la date, la filiation scientifique et académique. L'exercice proposé ci-dessous vise à repérer les différentes fonctions rhétoriques dans l'introduction.

**Extrait du corpus Scientext**

*L'analyse linguistique que nous proposons s'inscrit* en réponse à une demande de la SNCF relative à la perception du confort global en train. Les objectifs du projet sont d'identifier les conforts en train et leurs relations de dépendance, à partir d'analyses linguistiques et cognitives. *Cet article porte sur* la manière dont les formes linguistiques en contexte, utilisant les ressources de la langue mises en discours, renseignent sur les structures cognitives construites à partir des perceptions sensorielles. On *tente ici d'identifier*, à partir de l'analyse des formes adjectivales, *les représentations* individuelles et partagées qui se construisent dans les discours des voyageurs, lorsqu'on les interroge sur leur expérience sensible du confort. [...] *La première question que l'on se pose* alors est la suivante: le confort en train est-il une catégorie cognitive de ce type? [552-Communication-Linguistique-Introduction]

**Questions**

- *D'après vous, cet extrait vient de quelle partie de l'article? Et pourquoi?*
- *A. Résumé, B. introduction, C. théorie, D. méthodologie, E. conclusion.*
- *Découpez l'extrait en paragraphes et justifiez votre choix en précisant les motifs linguistiques et thématiques de votre découpage.*

Dans cet exercice, les apprenants doivent repérer les collocations (en italiques) qui les aident à découper cet extrait de l'introduction en quatre parties: contexte, sujet, objectif et question. Ce sont les principales fonctions rhétoriques dans l'introduction. L'enseignant peut également guider les apprenants vers d'autres constructions possibles pour le même sens.

Après le repérage, l'enseignant peut proposer un exercice de classement portant sur les fonctions rhétoriques étudiées. Il est conseillé de ne pas donner une liste récapitulative mais de la faire construire par les apprenants eux-mêmes. Ce travail de classement permet de fixer leurs connaissances et de créer un outil personnel pour leur rédaction. Par exemple, dans l'exercice ci-dessous, il est demandé aux apprenants de classer les expressions étudiées selon différentes fonctions rhétoriques venant de la partie méthodologique.

***Consigne**: Classez les expressions venant de la partie méthodologique et servant à présenter un choix de l'approche, la limite d'un choix et la collecte des données.*

| **Présenter un choix favorable** | **Présenter la limite d'un choix** | **Présenter la collecte des données** |
|---|---|---|
| *Nous adoptons l'approche […]* | *Nous nous limitons à étudier* | *Les données sont recueillies […]* |
| *Nous optons pour la méthode […]* | *On se restreint à l'analyse […]* | *rassembler des textes* |

La construction de la liste peut se fonder sur les exercices de repérage. Il s'agit donc d'un travail récapitulatif des collocations déjà rencontrées. Toutefois, l'enseignant peut également aider les apprenants à en trouver d'autres à l'aide du corpus. Cette liste peut varier en fonction de leurs acquis. Il est important de proposer une mise en commun des réponses pour s'assurer que les expressions erronées ne soient pas mémorisées.

En résumé, nous avons tenté de proposer quelques pistes visant à introduire le corpus en classe de FLE pour enseigner et apprendre les collocations. Les exercices

proposés ont été conçus dans l'objectif d'apporter aux étudiants différents contextes pour réutiliser les collocations et se familiariser avec leurs caractéristiques. Pour un apprentissage systématique de ces unités lexicales, il nous semble indispensable d'aborder les trois aspects lexico-syntaxique, syntagmatique et sémantico-discursif.

## Conclusion

Le corpus apporte de nombreux avantages pour l'apprenant et l'enseignant. Pour l'apprenant, le corpus est l'occasion d'accéder à de nombreux exemples. Il peut trouver seul des réponses et ainsi corriger ses propres erreurs en cherchant dans le corpus. Le corpus permettra donc à l'apprenant de développer sa production écrite. Pour l'enseignant, le corpus est un complément aux outils connus (dictionnaire, grammaire). Dans ce contexte, l'enseignant guide l'apprenant pour l'amener vers l'utilisation du corpus en autonomie.

Concernant la question principale de ce travail, comment enseigner les collocations à l'aide de corpus? Ces quelques pistes d'activités d'apprentissage ont été proposées en prenant en compte les propriétés des collocations. Les exemples de séquences autour des collocations sont loin d'être exhaustifs mais permettent de mieux comprendre le phénomène. Par ce travail, nous insistons particulièrement sur le rôle de l'enseignant dans le repérage des extraits de textes du corpus et dans la conception des activités. Celles-ci peuvent se trouver sous différentes formes (classification, reformulation, etc.) pour que l'apprenant puisse réutiliser les collocations dans un contexte universitaire.

## Références bibliographiques


AÏT-MOKHTAR, SALAH, JEAN-PIERRE CHANOD & CLAUDE ROUX. 2002. Robustness beyond shallowness: incremental dependency parsing. *Natural Language Engineering* 8: 2-3, 121-144.

BOLLY, CATHERINE. 2011. *Phraséologie et collocations. Approche sur corpus en français L1 et L2*. Bruxelles: PIE/ Peter Lang.

BOULTON, ALEX. 2008. Esprit de corpus: promouvoir l'exploitation de corpus en apprentissage des langues. *Texte et Corpus* 3, 37-46.

BOULTON, ALEX & HENRY TYNE. 2014. *Des documents authentiques aux corpus: démarches pour l'apprentissage des langues*. Paris: Didier.

CAVALLA, CRISTELLE. 2015. Collocations transdisciplinaires dans les écrits de doctorants FLS/FLE. *Linx* 72, 95-110.

CAVALLA, CRISTELLE & JULIE SORBA. 2017. Prendre un bain, des risques ou la fuite: étude diachronique du figement. FRANCIS GROSSMANN, SALAH MEJRI & INÈS SFAR (dir.). *Phraséologie: sémantique, syntaxe, discours*. Paris: Honoré Champion, 187-201.

CHAMBERS, ANGELA. 2005. Integrating corpus consultation in language studies. *Language Learning and Technology* 9, 111-125.

CHAMBERS, ANGELA. 2007. Popularising Corpus consultation by language learners and teachers, ENCARNACIÓN HIDALGO, LUIS QUEREDA & JUAN SANTANA (dir.). *Corpora in the foreign language classroom*. Amsterdam: Rodopi, 3-16.

EDMONDS, AMANDA. 2013. Une approche psycholinguistique des phénomènes phraséologiques: le cas des expressions conventionnelles. *Langages* 189, 121-138.

ELLIS, NICK C. 1997. Vocabulary acquisition: word structure, collocation, word-class, and meaning. NORBERT SCHMITT & MICHAEL MCCARTHY (dir.). *Vocabulary description, acquisition and pedagogy*. Cambridge: Cambridge University Press, 122-39.

FALAISE, ACHILLE. 2014. Exploitation linguistique de corpus arborés d'écrits scientifiques à l'aide du logiciel ScienQuest. AGNÈS TUTIN & FRANCIS GROSSMANN (dir.). *L'écrit scientifique: du lexique au discours. Autour de Scientext*. Rennes: Presses universitaires de Rennes, 123-143.

FIRTH, JOHN R. 1957. Modes of meaning. *Papers in linguistics 1934-1951*, 190-215.

FLIGELSTONE, STEVEN. 1993. Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME Journal* 17, 97-109.

GAVIOLI, LAURA. 2005. *Exploring corpora for ESP learning*. Amsterdam: John Benjamins.

GLEDHILL, CHRISTOPHER. 2011. The lexicogrammar approach to analysing phraseology and collocation in ESP texts. *ASp (Anglais de Spécialité)* 59, 5-23.

GROSSMANN, FRANCIS & AGNÈS TUTIN. 2003. *Les collocations: analyse et traitement*. Amsterdam: De Werelt.

HALLIDAY MICHAEL, ANGUS MCINTOSH & PETER STREVENS. 1964. *The linguistic sciences and language teaching*. London: Longman.

HAUSMANN, FRANZ JOSEF. 1985. Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. HENNING BERGENHOLTZ & JOACHIM MUGDAN (Hg.). *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch*, 118-129. Lexicographica. Series Maior 3.

HAUSMANN, FRANZ JOSEF. 1989. Le dictionnaire de collocations. FRANZ JOSEF HAUSMANN, OSKAR REICHMANN, HERBERT WIEGAND & LADISLAV ZGUSTA (dir.). *Wörterbücher: ein internationales Handbuch zur Lexicographie. Dictionaries. Dictionnaires*. Berlin/New-York: De Gruyter, 1010-19.

HILTON, HEATHER. 2005. Théories d'apprentissages et didactique des langues. *Les Langues Modernes* 3, 12-21. URL: http://cle.ens-lyon.fr/plurilangues/langue/didactique/theories-d-apprentissage-et-didactique-des-langues. Consulté le 20 mai 2018

JOHNS, TIM. 1991. Should you be persuaded: two examples of data-driven learning. Classroom concordancing. *English Language Research Journal* 4, 1-16.

JOHNS, TIM. 2002. Data-driven learning: the perpetual challenge. BERNHARD KETTEMANN & GEORG MARKO (dir.). *Teaching and learning by doing corpus analysis*. New York: Edwin Mellen, 107-118.

LARIVIÈRE, LOUISE. 1998. Valeur sémantique du verbe dans les collocations verbales spécialisées. *TTR: traduction, terminologie, rédaction* 11: 1, 173-197.

MÉDIONI, MARIA-ALICE. 2009. L'acquisition du vocabulaire: encore une question d'activité. *Les Langues Modernes* 1, 16-25.

MEL'CUK IGOR, ANDRÉ CLAS & ALAIN POLGUÈRE. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.

NESSELHAUF, NADJA. 2005. *Collocations in a learner corpus*. Amsterdam: John Benjamins.

PECMAN, MOJCA. 2005. Les apports possibles de la phraséologie à la didactique des langues étrangères. *Apprentissage des Langues et Systèmes d'Information et de Communication* 8: 1, 109-122.

POLGUÈRE, ALAIN. 2008. *Lexicologie et sémantique lexicale. Notions fondamentales*. Montréal: Presses de l'Université de Montréal.

SINCLAIR, JOHN MCHARDY. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

SINCLAIR, JOHN MCHARDY, SUSAN JONES & ROBERT DALEY. 2004. *English collocation studies: the OSTI report*. Londres & New York: Continuum.

TOGNINI-BONELLI, ELENA. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.

TRAN, THI THU HOAI. 2014. *Description de la phraséologie transdisciplinaire scientifique et réflexions didactiques pour l'enseignement à des étudiants non-natifs. Application aux marqueurs discursif*s. Thèse de doctorat, Université Grenoble Alpes.

TUTIN, AGNÈS. 2007. Collocations du lexique transdisciplinaire des écrits scientifiques: annotation et extraction des propriétés linguistiques dans la perspective d'une application didactique. *Cahiers de l'Institut de Linguistique de Louvain* 31: 2-4, 247-262.

TUTIN, AGNÈS. 2013. Les collocations lexicales: une relation essentiellement binaire définie par la relation prédicat-argument. *Langages* 1, 47-63.

TUTIN, AGNÈS & FRANCIS GROSSMANN. 2014. *L'écrit scientifique: du lexique au discours. Autour de Scientext*. Rennes: Presses universitaires de Rennes.

VAN DER LINDEN, ÉLISABETH. 2006. Lexique mental et apprentissage des mots. *Revue française de linguistique appliquée* 11: 1, 33-44.

WILLIAMS, GEOFFREY. 2006. La linguistique de corpus: une affaire prépositionnelle. FRANÇOIS RASTIER & MICHEL BALLABRIGA (dir.). *Corpus en lettres et sciences sociales: des documents numériques à l'interprétation*. Actes du colloque international d'Albi, 151-158. Paris: Texto.

YAN, RUI. 2017. *Étude des constructions verbales scientifiques dans une perspective didactique: utilisation des corpus dans le diagnostic des besoins langagiers du FLE à l'aide des techniques de TAL*. Thèse de doctorat en sciences du langage, Université Grenoble Alpes.

YAN, RUI & THI THU HOAI TRAN. 2018. *Les collocations génériques dans les écrits des étudiants allophones: diagnostic et remédiations*. Journée d'étude sur les collocations génériques, 15 janvier 2018, Université Paris Diderot – Paris 7.