

# Pour un enseignement systématique des marqueurs discursifs à l'aide de corpus en classe de FLE : l'exemple des marqueurs de reformulation

Thi Thu Hoai Tran, Agnès Tutin et Cristelle Cavalla (Grenoble)

---

## Abstract

This article focuses on the role of corpus in language classroom, a tool that allows to identify syntactic and semantic properties of discourse markers that are not discussed in traditional grammar books. The use of the corpus could assist French Foreign Language learners to be familiar with a new kind of academic writing. This raises questions about the methodology of teaching/learning of discourse markers that we try to address in this paper.

---

## 1 Introduction

Notre recherche s'inscrit dans la perspective de l'aide à l'écriture scientifique pour les étudiants non-natifs qui doivent rédiger un rapport de stage ou un mémoire en langue française. En effet, ces types d'écrits universitaires comportent des spécificités (structurelles et linguistiques) (Cavalla 2010: 154) et nécessitent d'être abordés en tant que tels. Notre intérêt porte essentiellement sur le lexique transdisciplinaire et spécifique dans les écrits scientifiques, et en particulier sur les séquences lexicalisées à fonction discursive (*dans un premier temps, au premier chef*, etc.) qui servent à structurer le discours et à organiser le texte. Ces unités lexicales revêtent différentes appellations comme *particules énonciatives* (Fernandez 1994), *connecteurs logiques* (souvent utilisés dans les méthodes de FLE), *discursive markers* en anglais ou *marqueurs discursifs* (désormais MD) en français (très en vogue actuellement).

En français, ces MD ont fait l'objet de nombreuses recherches qui proposent des traitements linguistiques assez fins parmi lesquelles *pourtant, pour autant* (Anscombe 1983), *en fait, de fait* (Danjou-Flaux 1980), les marqueurs de structuration de la conversation et de la complétude (Auchlin 1981), etc. Ces descriptions théoriques fort intéressantes sont cependant peu exploitées par les enseignants de FLE et encore moins par les apprenants. Reconnaissons que ce public n'a pas pour objectif d'entrer dans les distinctions fines de ces unités lexicales, mais les apprenants-étudiants doivent les connaître, car ils auront à les utiliser dans leurs écrits universitaires. Comme le public visé est constitué d'apprenants-étudiants de toutes les disciplines et que beaucoup ne sont pas linguistes, ils ne peuvent donc pas entrer dans des explications trop spécialisées ; c'est pourquoi, nous privilégions les explications non complexes au plan linguistique et donc didactisées.

L'objectif de cet article est double, théorique et didactique. Sur le plan théorique et linguistique, nous cherchons à comprendre le fonctionnement syntaxique et sémantique de certains MD. Sur le plan didactique, nous souhaitons mettre en place une méthodologie d'enseignement/apprentissage des MD, notamment en introduisant l'utilisation des corpus en classe de langue. En somme, nous envisageons de tirer profit des avantages de la linguistique de corpus pour la mise en place d'activités didactiques.

Cette recherche s'inscrit aussi dans la lignée d'autres travaux sur le lexique transdisciplinaire dans les écrits scientifiques, comme Phal (1971), Chetouani (1988), Coxhead (2000) d'une part, et Pecman (2004), Tutin (2007) d'autre part. Plusieurs typologies de ce lexique ont vu le jour et s'appuient sur des critères différents. Si Phal, Coxhead et Chetouani se fondent sur les critères lexicométriques pour créer des listes de mots pleins et mots grammaticaux, Pecman et Tutin se basent aussi sur les critères de combinatoire lexicale. Cette dernière méthode correspond mieux à notre perspective didactique, et nous la détaillerons dans la partie méthodologique.

Dans ce travail, nous proposons d'introduire le corpus en classe de langue pour aider les apprenants à s'approprier des phénomènes linguistiques spécifiques aux écrits scientifiques, les MD. L'utilisation de corpus dans l'enseignement/apprentissage des langues est très en vogue depuis les années 1990, particulièrement depuis les premiers travaux de Tim Johns (1991) sur le *Data-Driven Learning* (désormais DDL) ou « apprentissage guidé par les données » (traduction littérale). Selon Johns (1991), le DDL rend les apprenants plus autonomes dans leur apprentissage et leur permet de développer des stratégies de découverte en apprenant à apprendre ('learn to learn') (*ibid.*: 1). Nous reprenons de notre côté la conception plus large de Boulton et Tyne (2014) de l'*apprentissage sur corpus* (désormais ASC) (*ibid.*), qui renvoie à « toute utilisation de données tirées de façon explicite d'un corpus pour exploitation par des non natifs à des fins d'apprentissage ou d'aide à l'expression écrite, la traduction, etc ». (*ibid.*: 145). La notion d'ASC correspond bien à notre démarche quand nous envisageons de développer la réflexion inductive des apprenants tout en exploitant des éléments linguistiques.

Cependant, l'introduction du corpus en classe de FLE ne va pas de soi. Bien que cette méthode apparaisse intéressante et efficace pour la prise de conscience du phénomène linguistique (cf. expérimentations de Landure 2013), comme souligné par Cavalla et Loiseau (2013: 175), les enseignants de FLE connaissent encore peu les corpus et leur utilité, et restent peu formés à leur utilisation. Différents travaux tentent d'expliquer cet état de fait par des difficultés cognitives des apprenants, la nature de l'ASC, l'absence de formation des enseignants de FLE. Parmi les difficultés rencontrées, Gilquin et Granger (2010: 366) ont mis en évidence la difficulté cognitive de demander aux apprenants de formuler par eux-mêmes les règles. Le rôle de l'enseignant apparaît indispensable pour orienter l'apprentissage et les aider à trouver les propriétés syntaxiques et sémantiques des éléments linguistiques. Cette difficulté est aussi souvent liée au nombre important de résultats obtenus dans les requêtes sur corpus. Comme Gilquin et Granger (2010: 366), nous pensons que les apprenants risquent de se perdre face à un nombre d'occurrences élevé. C'est pourquoi, dans cette recherche, nous tentons de mettre en avant la sélection préalable des exemples extraits du corpus par l'enseignant de façon que les apprenants puissent se concentrer sur un phénomène linguistique précis.

L'absence d'utilisation du corpus en classe de FLE est largement liée aux pratiques des enseignants. D'une part, ceux-ci sont parfois réticents à modifier leurs pratiques d'enseignement avec l'introduction du corpus en classe. D'autre part, Gilquin et Granger (2010: 366) ont souligné un manque de connaissances sur les possibilités offertes par les corpus dans l'enseignement/apprentissage. À l'instar de Mauranen (2004), nous insistons sur l'importance de la formation et de la prise de conscience des enseignants face à la richesse du corpus pour l'élaboration du contenu d'enseignement. Ils doivent être les premiers à reconnaître les avantages du corpus afin d'expliquer aux étudiants les phénomènes linguistiques.

Face à l'introduction du corpus en classe de FLE, nous allons tenter de répondre à deux questions :

- 1) Quels sont les éléments à prendre en compte pour aborder les MD ?
- 2) Quelles sont les méthodes pour introduire le corpus en classe de langue ?

Cet article s'organise de la manière suivante. Après l'étude de quelques grammaires de FLE, nous effectuerons l'analyse linguistique d'un des marqueurs de reformulation les plus fréquents dans notre corpus, le marqueur *c'est-à-dire*. Cette analyse nous servira d'appui par la suite pour des suggestions didactiques dans la deuxième partie.

## 2 Intérêts des corpus pour l'enseignement/apprentissage du français

Dans cette partie, nous procédons à l'étude des livres de grammaire de FLE afin de comprendre comment les MD sont introduits. Face au nombre très important d'ouvrages de ce type, nous avons choisi les grammaires utilisées dans les filières de l'Agence Universitaire de la Francophonie (AUF) au Vietnam et au Département de français, à l'Université nationale de Hanoï (Tableau 1).

Grammaire	Editeur	Editeur scientifique	Année de parution
La grammaire des premiers temps (niveau 2)	PUG	Dominique Abry Marie-Laure Chalaron	1999
Grammaire du français	Hachette	Yves Delatour	1991
Grammaire utile du français	Didier	Evelyne Berard Christian Lavenne	1991
Précis de grammaire	CLE International	Isabelle Chollet Jean-Michel Robert	2005
Grammaire expliquée du français (Niveau intermédiaire)	CLE International	Sylvie Poisson-Quinton Reigne Mimran Michèle Mahéro-Le Coadic	2002
Grammaire pour l'enseignement/apprentissage du FLE	Didier / Hatier	Geneviève-Dominique de Salins	1996

Tableau 1 : Livres de grammaire de FLE consultés

La plupart des grammaires choisies sont des grammaires pour l'enseignement, sauf *La grammaire pour l'enseignement/apprentissage du FLE* qui peut être considérée comme une grammaire « de transition », c'est-à-dire une grammaire pour la formation des enseignants.

En effet, le public visé est celui de jeunes enseignants prêts à entamer une réflexion sur l'utilisation des notions grammaticales. Cette démarche correspond à notre perspective, car les réflexions didactiques proposées dans la dernière partie s'adressent en premier lieu aux enseignants.

De cette brève analyse, ressortent un manque d'explications précises de l'utilisation des MD dans certaines grammaires et l'absence d'exemples authentiques. Les MD sont généralement abordés très brièvement ou sous la forme de listes de mots. Dans la *Grammaire des premiers temps*, ou dans la *Grammaire expliquée du français*, chaque MD est décrit brièvement en une ou deux phrases, les phrases – énoncés exemplaires – tenant lieu d'explication. Certes, ces brèves définitions peuvent tout à fait répondre aux besoins d'un public non universitaire qui souhaite connaître l'emploi général du marqueur, mais un public universitaire aura besoin, à notre avis, d'explications plus détaillées afin de distinguer un MD d'un autre dans ses futurs écrits universitaires. Il est indispensable d'expliquer aux apprenants à la fois les propriétés syntaxiques et sémantiques de ces unités lexicales.

Par ailleurs, les exemples présentés apparaissent hors contexte, tandis que nos publics universitaires doivent répondre aux normes conventionnelles d'un genre spécifique. Des exemples authentiques extraits des textes scientifiques leur permettraient, non seulement de s'entraîner dans un contexte proche des écrits attendus, mais aussi de se familiariser au langage d'un genre spécifique comme les écrits scientifiques.

En bref, ce petit diagnostic dans les grammaires d'enseignement nous apporte une vue globale sur la manière dont les MD sont abordés et montre l'intérêt d'une approche plus systématique recourant à des exemples authentiques. Nous souhaiterions montrer que le travail sur corpus permet d'éviter de recourir aux explications linguistiques complexes. Nous allons maintenant décrire notre corpus de travail et l'approche privilégiée pour l'aborder.

### **3 Méthodologie**

#### **3.1 Corpus et extraction des MD**

Le corpus constitué doit répondre avant tout aux objectifs didactiques. Comme nous envisageons de proposer une aide à l'écriture universitaire à destination des étudiants en Sciences Humaines et Sociales (SHS), ce corpus doit à la fois répondre à une certaine diversité disciplinaire et à une homogénéité dans le genre. Notre corpus se situe dans la continuité du corpus Scientext<sup>1</sup> et a bénéficié des travaux antérieurs, particulièrement en ce qui concerne les modes d'interrogation du corpus. Notre corpus recense environ 2 680 000 mots et se compose d'articles scientifiques issus de 10 disciplines en SHS (30 articles par discipline) : linguistique, psychologie, sciences de l'éducation, économie, histoire, géographie, sociologie, anthropologie, sciences politiques, sciences de l'information et de la communication.

Ce corpus a été étiqueté morpho-syntaxiquement et annoté automatiquement grâce à l'analyseur syntaxique Syntex de Bourigault (2007). Nous avons extrait les séquences

---

<sup>1</sup> Le projet ANR Scientext (2007-2010) été piloté par Agnès Tutin et Francis Grossmann du LIDILEM, le site est consultable à l'adresse suivante: <http://scientext.msh-alpes.fr/scientext-site/spip.php?article1> [30.06.2015].

lexicalisées à fonction discursive à l'aide de n-grammes (Granger/Paquot 2008) et en extrayant les locutions adverbiales et prépositionnelles du corpus. Nous avons retenu les séquences à la fois transdisciplinaires (au moins 4 disciplines sur 10) et fréquentes (au moins 7 occurrences).

Nous avons relevé dans le corpus 171 séquences lexicalisées à fonction discursive (Tran 2014) qui sont essentiellement des locutions adverbiales, mais aussi des locutions prépositionnelles, conjonctives et des structures phrastiques. À l'instar de Burger (1998) et Tutin (2014), nous nous appuyons sur les fonctions que jouent ces MD dans l'énoncé et en distinguons deux fonctions principales à savoir les marqueurs à fonction métadiscursive et ceux qui ont une fonction argumentative. Dans notre typologie, les marqueurs à fonction métadiscursive englobent toutes les locutions qui jouent un rôle de guidage du locuteur. Il peut s'agir des marqueurs d'énumération (*tout d'abord*) et de topicalisation (*à ce propos*) qui introduisent une nouvelle thématique ou un sujet, ou bien des marqueurs de reformulation (*à savoir, c'est-à-dire*) qui permettent au scripteur de clarifier son énoncé. Les marqueurs à fonction argumentative comprennent les locutions souvent appelées connecteurs logiques comme les marqueurs de concession (*tout de même*), d'opposition (*à l'encontre de*) ou de finalité (*en vue de*), etc.

Dans ce travail, notre intérêt porte sur des marqueurs de reformulation très fréquents dans les écrits scientifiques. Nous insistons particulièrement sur l'avantage de l'apprentissage de ces marqueurs de reformulation pour la production écrite des étudiants. L'utilisation ces marqueurs permet aux étudiants à la fois de mieux structurer leur écrit et d'éviter la juxtaposition des éléments et d'explicitier le message. Nous nous inspirons des travaux de Roulet (1987) et de Rossari (1994) et distinguons trois fonctions principales des marqueurs de reformulation : 1) la **clarification** (*c'est-à-dire, à savoir, autrement dit, en d'autres termes*), où le scripteur revient sur ce qui a été dit dans l'énoncé précédent afin de l'éclairer par des termes connus par son interlocuteur ; 2) la **récapitulation** (*en somme, en résumé*) qui sert à reprendre les idées principales ; 3) et la **reconsidération** (*en définitive, au fond*) qui permet d'introduire le point de vue de l'auteur.

L'analyse des MD s'effectue selon une approche empirique, mais aussi comparée, afin de différencier un MD d'un autre à partir des propriétés syntaxiques et sémantiques observées. Dans le cadre de cet article, nous allons analyser un marqueur de reformulation particulièrement fréquent dans les écrits scientifiques, *c'est-à-dire*.

### 3.2 Grille d'analyse

Le choix des paramètres à prendre en compte pour l'analyse syntactico-sémantique n'est pas simple en soi. Les paramètres retenus doivent être clairs et facilement exploitables dans le cadre de séquences didactiques. Pour cette recherche, nous décrivons les MD à l'aide des trois paramètres suivants : la *portée syntaxique*, la *position syntaxique* et les *valeurs sémantiques*.

Premièrement, la *portée syntaxique* représente pour nous une notion très féconde qui a fait l'objet de nombreuses recherches, par exemple Muller (1991, 2005), Guimier (1996), Charolles et Péry-Woodley (2005). Comme Rossari et Jayez (1997: 234), nous pensons que la portée permet de faire ressortir les autres propriétés syntactico-sémantiques des MD. La portée du MD est déterminée, selon nous, par son fonctionnement sémantique. À l'instar de

Paillard et Vu (2012), nous distinguons la *portée globale* de la *portée locale*. Si le MD porte sur une proposition ou un énoncé, il a une *portée globale* (**P<sub>G</sub>**), alors que si le MD porte sur un syntagme, il a une portée locale (**P<sub>L</sub>**). L'observation du rapport entre le contexte gauche et la portée du MD révèle des propriétés syntaxiques intéressantes. Nous allons y revenir dans l'analyse linguistique de *c'est-à-dire* ci-dessous.

Deuxièmement, nous nous intéressons également à la *position syntaxique* des MD. Comme nous l'avons évoqué, notre typologie porte essentiellement sur des locutions adverbiales, qui peuvent se trouver en position initiale, médiane ou finale. Pourtant, les différentes positions phrastiques des adverbiaux ne sont pas aléatoires et liées aux objectifs communicatifs de l'auteur comme l'a montré Charolles (1997, 2003).

Troisièmement, l'étude de la *portée syntaxique* des MD et de leur *position syntaxique* sert à relever leurs *valeurs sémantiques* qui peuvent être multiples. Nous attribuons une étiquette pour chaque valeur sémantique correspondante. Cette tâche n'est pas simple, mais apparaît indispensable pour un travail de didactisation, par exemple, pour une approche onomasiologique qui propose un accès par le sens des MD.

Nous nous inspirons dans ce travail du modèle proposé par Paillard et Vu (2012). Les séquences lexicalisées à fonction discursive seront analysées selon le schéma suivant :

#### **p MD q**

**MD**: marqueurs discursifs, qu'il s'agisse des monolexicaux ou les polylexicaux ;

**p**: la séquence qui précède le MD ;

**q**: la séquence qui suit le MD et correspond à sa portée.

#### **4 Analyse syntactico-sémantique de *c'est-à-dire***

Nous allons appliquer le modèle présenté ci-dessus à l'analyse de *c'est-à-dire*, le marqueur de reformulation le plus utilisé dans notre corpus, au total 704 occurrences réparties dans 10 disciplines. Force est de constater que ce marqueur a fait l'objet de nombreuses études (Steuckardt et Niklas-Salmien 2005 ; Vassiliadou 2008 ; Khatchatourian 2008). Ces descriptions très détaillées sur le plan énonciatif ou diachronique contribuent à mieux comprendre les valeurs sémantiques de ce marqueur.

Khatchatourian (2008) propose de distinguer les différents emplois des marqueurs *c'est-à-dire* et *cela veut dire* en partant du sémantisme du verbe *dire* en s'appuyant sur les problèmes principaux de la communication. Vassiliadou (2008) effectue une analyse contrastée sur quatre marqueurs de reformulation *c'est-à-dire*, *à savoir*, *autrement dit*, *en d'autres termes*. Cette étude permet de relever les rapports entre ces marqueurs et également de mettre en évidence leurs différentes contraintes d'emploi. Les différences sont dues, d'une part, au sémantisme de chaque marqueur, et d'autre part, à la structure syntaxique. Les rapports établis entre les énoncés grâce aux marqueurs de reformulation **p** et **q** figurent sur deux plans, une équivalence métalinguistique et une équivalence qui relève de divers mécanismes comme des équivalences par métaphores, métonymies, synecdoques. Outre ces régularités, Vassiliadou (*ibid.*) distingue l'emploi de ces marqueurs par des tests de compatibilité en prenant comme référence le marqueur *c'est-à-dire* – le marqueur de reformulation le plus prototypique selon

elle. Ces tests révèlent des cas où les quatre marqueurs peuvent commuter, ainsi que des emplois qui spécifient un marqueur parmi d'autres.

Comme nous l'avons précisé, les descriptions linguistiques que nous allons effectuer ci-dessous seront adaptées à une utilisation didactique. L'intérêt de cette recherche est d'évaluer les propriétés des MD dans les écrits scientifiques à la lumière du corpus.

#### 4.1 Portée syntaxique

Si l'on reprend la définition du TLFi<sup>2</sup>, le rôle que joue ce MD dans l'énoncé est d'introduire une information complémentaire ou rectificative. Nous distinguons *c'est-à-dire* en tant que locution adverbiale de *c'est-à-dire que* comme locution conjonctive. En ce qui concerne la portée<sup>3</sup> de ces marqueurs, *c'est-à-dire* a en général une portée locale (**PL**), quand il porte sur un syntagme adjectival 1), un syntagme nominal, un syntagme verbal, ou un syntagme prépositionnel. Pourtant, ce MD peut également porter sur une proposition et avoir dans ce cas une portée globale (mais cet usage est moins fréquent). *C'est-à-dire que*, en raison de sa nature syntaxique porte toujours sur une proposition 2) et a une portée globale (**PG**). Il est beaucoup moins fréquent que *c'est-à-dire* (57 occurrences).

1) Tertullien ajoute que le vêtement « est entré » dans le Latium avec le mot, indiquant par là [l'origine étrangère] (**c'est-à-dire grecque**) du pallium, sans aucune précision chronologique. [his-art-127-body]

2) Comme le souligne Passeron, [« discours statistique et le discours sociologique différent intrinsèquement par leur pertinence empirique], **c'est-à-dire qu'ils ne sont pas justiciables, pour trancher de la vérité ou de la fausseté de leurs énoncés, de procédures de contrôle référant leurs expériences de dernière instance à la même réalité empirique** » 1991: 113). [sed-art-8-notes]

En observant la portée de *c'est-à-dire* et son contexte gauche, nous constatons qu'il existe une certaine équivalence de la nature syntaxique entre **p** et **q** ; lorsque le contexte gauche **p** est un syntagme nominal ou verbal ou une proposition, la séquence **q** (dans *c'est-à-dire q*) est de même nature.

#### 4.2 Position syntaxique

La position de *c'est-à-dire* et de *c'est-à-dire que* est liée à leur portée. Ces MD se trouvent toujours en position initiale et sont suivis d'une séquence reformulatrice (2), (3). *C'est-à-dire* ne peut ni se trouver en position médiane ni en position finale<sup>4</sup>.

---

<sup>2</sup> Trésor de la Langue Française informatisé consultable à l'adresse suivante : [www.atilf.fr/spip.php?rubrique77](http://www.atilf.fr/spip.php?rubrique77).

<sup>3</sup> Nous avons souligné la portée de *c'est-à-dire* dans les exemples extraits du corpus et la partie gauche du MD se trouve entre crochets.

<sup>4</sup> Dans notre typologie des séquences lexicalisées à fonction discursive (Tran 2014), *en somme* fait partie des marqueurs de reformulation qui servent à récapituler un énoncé. Pourtant, quand ce MD se trouve en position finale, il exprime plutôt une fonction modale que de reformuler l'énoncé précédent comme dans l'exemple suivant : « Mais cette population est terriblement fragilisée par la stagnation de son pouvoir d'achat face à l'augmentation du foncier : la plupart de ces habitants ont contracté des emprunts auprès de leurs interlocuteurs

- 3) La répétition de ce rituel produit une certaine confiance parce qu'il témoigne d'une certaine stabilité : elle inscrit cette rencontre avec l'autre à peine identifié dans les limites d'un monde connu : elle contribue à ce que Giddens (1987) appelle [la « sécurité ontologique »], **c'est-à-dire** le sentiment que le monde est ordonné et stable, qu'il va de soi. [soc-art-59-body]

Ajoutons que pour l'enseignement/apprentissage des MD, il semble important d'aborder les régularités typographiques des MD<sup>5</sup>, car les étudiants viennent peut-être de traditions graphiques très éloignées de la langue française. La typographie est l'une des sources d'erreurs fréquentes chez les étudiants non-natifs et également des natifs du français. *C'est-à-dire* est en général précédé d'une virgule (608 sur le total de 704 occurrences de *c'est-à-dire* dans notre corpus) mais nous avons relevé 36 occurrences sans ponctuation.

### 4.3 Valeurs sémantiques

Nous distinguons quatre valeurs sémantiques principales de *c'est-à-dire* : la **valeur synonymique**, la **valeur interprétative**, la **valeur énumérative** et la **valeur résumante**.

#### 4.3.1 Valeur synonymique

*C'est-à-dire* peut introduire une séquence à **valeur synonymique** quand il établit un rapport synonymique entre **p** et *c'est-à-dire* **q**. Ce rapport se construit par des procédés de définition (4), (5). Il peut s'agir d'une définition classique (4) ou bien d'une définition choisie par l'auteur qui correspond au contexte d'énonciation (5).

- 4) Selon E.C. Hugues (Hugues E.C. 1996) une profession est caractérisée par le fait que l'exercice de ses activités nécessite l'obtention d'une [licence], **c'est-à-dire** d'une permission légale d'exercer un certain travail interdit aux autres individus qui n'en sont pas détenteurs (ce qui peut correspondre, pour ce qui concerne les enseignants par exemple, à l'obtention du CRPE, CAPES, CAPET, CAPLP, agrégation, etc. [sed-art-124-body]

- 5) Ce type de configuration spatiale augmente la probabilité d'un [effet domino], **c'est-à-dire** « la survenance en cascades de dommages et de perturbations suite à un premier choc, au sein de l'enjeu touché, mais aussi dans son environnement (entendu au sens large d'environnement socio-économique) » (Provitolo 2007 ; Gleyze/Reghezza 2007). [geo-art-111-body]

Dans l'exemple (4), **q** (dans *c'est-à-dire* **q**) est seulement une des définitions normées du mot *licence*. Le terme *licence* peut référer soit au diplôme national sanctionnant les trois premières années d'études du premier cycle de l'enseignement supérieur, soit au document administratif attaché à un véhicule routier, soit au permis d'exercer une activité, etc. Le marqueur *c'est-à-dire* introduit dans ce cas la définition du terme *licence* dans ce contexte et permet au locuteur de délimiter le référent de la *licence* et de faciliter le travail d'interprétation du lecteur.

Dans (5), nous supposons que *effet domino* peut avoir différentes définitions. Pourtant, dans le contexte de son travail, l'auteur a adopté uniquement une définition précise correspondant à cette citation.

---

immobiliers mafieux, avec des clauses de réajustement de la dette en fonction de l'évolution du coût de la pierre... Une « gentrification » nourrit l'autre, en somme ». [soc-art-458-body]

<sup>5</sup> Nous tenons à remercier Alain Kamber de ses remarques pertinentes sur ce point.



### 4.3.2 Valeur interprétative

*C'est-à-dire* peut avoir une **valeur interprétative** quand la reformulation établie par *c'est-à-dire* est employée pour interpréter l'énoncé **p** (6), mais ici, la séquence **q** (*c'est-à-dire q*) correspond à une conception subjective du locuteur de l'état de choses, contrairement à la valeur synonymique. *C'est-à-dire* précise le point de vue du locuteur, et apporte également d'autres informations qui permettent de situer le contexte (6).

- 6) À titre d'exemple, la rubrique d'un petit forum d'initiés concernant la série Breaking Bad, forte de nombreux échanges entre une douzaine de fans durant la deuxième saison présentée aux États-Unis, cesse brusquement toute activité pendant l'intersaison, de juin 2009 à mars 2010, pour reprendre précisément [le 22 mars], **c'est-à-dire le jour suivant la diffusion américaine du premier épisode de la troisième saison.** [infcom-art-21-body]

Dans l'exemple (6), l'utilisation de ce marqueur permet à l'auteur de mettre l'accent sur l'évènement afin d'éclairer le lecteur sur ce qui s'est passé dans un contexte précis. *C'est-à-dire* peut aussi renvoyer directement à un élément évoqué précédemment dans le texte 7.

- 7) En ce qui concerne [la solution (3)], **c'est-à-dire l'émission de certificats de dépôts**, il s'agit de savoir [si la demande de certificats de dépôts est parfaitement élastique au taux du marché], **c'est-à-dire si les banques peuvent émettre autant de certificats de dépôts qu'elles le veulent sans payer de prime.** [eco-art-572-body]

Dans l'exemple (7), *c'est-à-dire* est utilisé deux fois dans la même phrase. Si le premier *c'est-à-dire* sert à référer à ce qui a été dit, il correspond à un rappel d'analyses antérieures. Le deuxième *c'est-à-dire* a une valeur interprétative quand l'auteur cherche à expliciter ce qu'il entend par l'élasticité au taux du marché de la *demande de certificats de dépôts*.

### 4.3.3 Valeur énumérative

*C'est-à-dire* peut aussi avoir une **valeur énumérative** (8), lorsqu'il sert à introduire une énumération. Dans l'exemple (8), l'auteur énumère dans la séquence **q** les différents *experts investis dans les organismes nationaux*.

- 8) Son effectif est constitué [des experts déjà investis dans les organismes nationaux], **c'est-à-dire des agronomes et des juristes qui se réclament de travaux publiés par les géographes.** [spo-art-51-body]

*C'est-à-dire* y instaure en quelque sorte une relation d'hyponymie en allant de quelque chose plus général (dans **p**) à quelque chose de plus spécifique ou plus précis (dans **q**).

### 4.3.4 Valeur résumante

*C'est-à-dire* peut aussi introduire une séquence à valeur résumante, même si cet usage est peu fréquent dans notre corpus (9).

- 9) Or en focalisant l'attention sur les techniques gestionnaires, la « nouvelle gestion publique » élude les débats sur [les projets collectifs et les arbitrages], **c'est-à-dire le politique** (Vinokur 2008). [soc-art-377-body]

Dans cet exemple, la séquence reformulatrice **q** correspond à un condensé de ce qui a été dit dans **p** sur *les projets collectifs et les arbitrages*.

#### 4.4 Synthèse

En comparant la taille de la portée **q** de *c'est-à-dire* à celle de son contexte gauche **p**, nous constatons des régularités sémantiques, la taille de **p** et **q** étant définie par la délimitation graphique de ces deux séquences. Si la taille de **p** et **q** est équivalente, *c'est-à-dire* sert en général à introduire des définitions, des explications, etc., et il a une **valeur synonymique** ou **interprétative**. Par contre, si la taille de **p** est inférieure à **q**, **q** correspond à une énumération, *c'est-à-dire* introduit une séquence à **valeur énumérative**. Enfin, quand la taille de **p** est supérieure à **q**, **q** est la séquence synthétique de **p**, elle a une **valeur résumante**.

Afin de faciliter la mémorisation et les utilisations didactiques, les propriétés syntaxiques et sémantiques de *c'est-à-dire* peuvent être présentées sous forme de tableau tel le Tableau 2 ci-dessous.

Propriétés syntactico-sémantiques	C'est-à-dire (que)
<b>Portée syntaxique</b>	<p><u>Locale (PL)</u> :</p> <p>[syntagme nominal]<sub>p</sub>, <b>c'est-à-dire</b> [syntagme nominal]<sub>q</sub></p> <p>[syntagme verbal]<sub>p</sub>, <b>c'est-à-dire</b> [syntagme verbal]<sub>q</sub></p> <p>[syntagme adjectival]<sub>p</sub>, <b>c'est-à-dire</b> [syntagme adjectival]<sub>q</sub></p> <p>[syntagme prépositionnel]<sub>p</sub>, <b>c'est-à-dire</b> [syntagme prépositionnel]<sub>q</sub></p> <p><u>Globale (PG)</u> uniquement pour <i>c'est-à-dire que</i> :</p> <p>[proposition]<sub>p</sub>, <b>c'est-à-dire que</b> [proposition]<sub>q</sub></p>
<b>Position syntaxique</b>	<p><u>Initiale</u> : <i>C'est-à-dire</i> introduit la séquence reformulatrice, il est placé en position initiale. Au niveau phrastique, il se trouve toujours en position post-verbale.</p> <p>[ [ ]<sub>p</sub>, <b>c'est-à-dire (que)</b> [ ]<sub>q</sub> ]<sub>phrase</sub></p>
<b>Trait typographique</b>	<i>C'est-à-dire</i> et <i>c'est-à-dire que</i> sont précédés d'une virgule.
<b>Valeurs sémantiques</b>	<p><b>Valeur synonymique :</b></p> <p>Le MD introduit une définition, une dénomination, une explication qui est construite par un rapport synonymique.</p> <p>[ ]<sub>p</sub> <b>c'est-à-dire (que)</b>  V.SYNONYMIQUE  [ ]<sub>q</sub> (<b>p = q</b>)</p> <p><b>Valeur interprétative :</b></p> <p>Le MD apporte de propres interprétations du locuteur (par des conceptions des choses, par renvoi dans le texte, etc.).</p> <p>[ ]<sub>p</sub> <b>c'est-à-dire (que)</b>  V.INTERPRETATIVE  [ ]<sub>q</sub> (<b>p = q</b>)</p> <p><b>Valeur énumérative :</b></p> <p><i>C'est-à-dire</i> introduit une énumération. L'utilisation du</p>

	<p>déterminant défini indique que l'information est connue par le lecteur par ses propres connaissances ou par les informations dans le texte.</p> <p>[ ]<sub>p</sub> <b>c'est-à-dire</b>  V.ENUMERATIVE  [ ]<sub>q</sub> (<b>p &lt; q</b>)</p> <p><b>Valeur résumante :</b></p> <p><i>C'est-à-dire</i> introduit une idée synthétique.</p> <p>[ ]<sub>p</sub> <b>c'est-à-dire (que)</b>  V.RESUMANTE  [ ]<sub>q</sub> (<b>p &gt; q</b>)</p>
--	--

**Tableau 2 : Récapitulatif des propriétés syntaxiques et sémantiques de *c'est-à-dire***

Cette analyse du marqueur *c'est-à-dire* montre l'intérêt d'une étude sur corpus. Celle-ci sert à relever les propriétés saillantes de chaque MD et à mettre en relief les éléments à aborder lors de l'enseignement de ces unités lexicales. Ces analyses nous apportent des données riches pour proposer une méthodologie adaptable à l'enseignement/apprentissage des MD au moyen du corpus. Dans la partie didactique, nous allons nous baser sur ces descriptions linguistiques pour présenter quelques réflexions didactiques dans l'objectif d'initier les étudiants aux caractéristiques des écrits scientifiques.

## 5 Propositions didactiques

Nous privilégions dans ce travail l'approche inductive pour aider les étudiants à travailler sur des phénomènes linguistiques à partir du corpus. Cette approche, qui décrit un type de raisonnement qui va du particulier au général, est souvent introduite dans l'enseignement des phénomènes grammaticaux. Dans le domaine de la linguistique de corpus, Chambers (2010), Boulton et Tyne (2014) y recourent pour l'enseignement/apprentissage.

Notons que nous mettons principalement l'accent sur les compétences de la compréhension écrite et de la production écrite. Dans les lignes qui suivent, nous allons proposer quelques activités basées sur corpus qui permettent aux étudiants à la fois de travailler sur les MD et de se familiariser aux propriétés des écrits scientifiques. Nous distinguons deux étapes d'un travail sur corpus : une observation des concordances des MD et une observation des MD dans un contexte large. Ces activités d'observation pourront être complétées par une production écrite.

### 5.1 Observation des concordances

Nous allons proposer ici un exemple d'utilisation du corpus en classe de langue pour enseigner les MD, plus précisément le marqueur *c'est-à-dire*. Pour enseigner un phénomène linguistique comme *c'est-à-dire*, l'enseignant peut constituer de petits sous-corpus extraits du concordancier de ce MD dans le corpus Scientext<sup>6</sup>. Notre public est composé d'étudiants au niveau B2 du CECRL. L'analyse d'un corpus composé de mémoires d'apprenants vietnamophones (Tran 2014) nous permet de constater que les étudiants de ce niveau

<sup>6</sup> La question du choix des textes (même discipline que celle des étudiants ou pas) fait débat mais nous ne développerons pas ce point ici.

connaissent *c'est-à-dire*, mais le confondent avec *à savoir*. Cet exercice peut être considéré comme une première entrée dans les corpus.

*C'est-à-dire* fait partie des MD les plus productifs de notre corpus (704 occurrences). Comme nous ne pouvons pas faire travailler les apprenants sur toutes les occurrences de ce marqueur pour faire découvrir ses différentes valeurs, l'enseignant sélectionne préalablement quelques occurrences où *c'est-à-dire* a la même valeur d'usage (il s'agit ici de l'approche onomasiologique) (Figure 1). Dans cet exercice, *c'est-à-dire* introduit une valeur synonymique (Tableau 2). Cette concordance peut être utilisée comme un exercice d'observation dans une démarche inductive. L'enseignant fait repérer le sens de *c'est-à-dire* dans ces exemples, en demandant aux apprenants d'analyser la structure syntaxique du contexte gauche et du contexte droit de *c'est-à-dire*. Ensuite, l'enseignant demande de relever le rapport entre les deux contextes et formuler le rôle syntaxique et sémantique que joue le MD.

À condition toutefois de désenclaver la réception, <i>c'est-à-dire</i> de définir celle-ci comme une activité socialisée et nomade, indépendamment du caractère éventuellement résidentiel des objets techniques. [infcom-art-5-body]
Il s'ensuit que les séries télévisées constituent des vecteurs majeurs pour connaître la police ; il s'agit d'un savoir collectif acquis par procuration, <i>c'est-à-dire</i> par médias interposés. [infcom-art-5-body]
Ce qui explique, par exemple, que la principale forme de judiciarisation puisse être le « <i>judicialreview</i> » <i>c'est-à-dire</i> le contrôle des lois par les tribunaux (Vallinder, 1994 ; Hertog et Halliday, 2004) ou, plus précisément encore, le « <i>constitutionalreview</i> » (Shapiro et Stone, 2004). [soc-art-35-body]
D'une part, <i>c'est</i> la seule qui marque sa singularité en s'appuyant sur les conduites motrices des élèves ; d'autre part, <i>c'est</i> également la seule qui est confondue avec les supports qu'elle emploie, <i>c'est-à-dire</i> avec les activités physiques et sportives. [sed-art-52-body]

Figure 1 : Concordance de *c'est-à-dire*

L'enseignant peut intervenir pour guider les apprenants à repérer les propriétés syntaxiques et sémantiques de ce MD. L'enseignant se sert du Tableau 2 afin d'obtenir les idées essentielles à aborder pour les apprenants. *C'est-à-dire* peut porter sur un groupe nominal, un groupe verbal, un groupe prépositionnel, ou une proposition. Quand *c'est-à-dire* porte sur une proposition, il est suivi de *que*. L'apprentissage évolue dans une approche inductive guidée.

Afin d'aider les apprenants à acquérir les propriétés spécifiques de chaque MD, l'enseignant pourra proposer des exercices dans lesquels des substitutions des MD sont possibles, comme *c'est-à-dire* avec d'autres marqueurs de reformulation *à savoir*, *autrement dit*, *en d'autres termes* afin de mieux montrer dans quels cas utiliser tel MD. Ces activités permettent selon nous d'éviter le recours automatique à la notion de *synonymie* dans l'enseignement du lexique.

## 5.2 Observation des MD dans les textes

Sachant que l'utilisation du corpus ne s'arrête pas au concordancier, nous envisageons de faire travailler les apprenants sur un contexte plus large. L'objectif est qu'ils prennent conscience de la manière dont ces éléments linguistiques participent à la structuration du texte. L'enseignement ne se limite pas au niveau lexical, mais doit tenir compte d'autres aspects des écrits scientifiques, à savoir structurel et pragmatique, car l'objectif final est d'aider les apprenants à s'appropriier les codes propres à ce genre d'écrits.

Dans ce travail, nous souhaitons développer chez les apprenants des stratégies de lecture pour développer par la suite les stratégies de production. Les apprenants seront amenés à observer

des modèles de structuration du texte-exemple pour élaborer un plan de leur écrit. En fait, l'analyse du corpus et les activités de compréhension écrite apportent des ressources lexicales et des savoirs langagiers indispensables à la production écrite.

Comme Lundquist (2013), nous pensons qu'il est important d'enseigner aux apprenants une méthodologie de travail afin qu'ils deviennent autonomes dans leur apprentissage. Nous nous inspirons de deux processus proposés par Lundquist (*ibid.*): le *top-down* (du haut vers le bas) et le *bottom-up* (du bas vers le haut). La démarche *top-down* concerne les connaissances sur les structures discursives dans l'écrit scientifique afin de faire ressortir les éléments sémantiques et structurels, tandis que la stratégie de lecture *bottom-up* part des éléments linguistiques (lexique et syntaxe) pour construire la compréhension du texte. Prenons l'exemple d'une introduction extraite d'un article scientifique (Figure 2), partie textuelle stratégique dans ce type d'écrits. On y relève essentiellement les éléments suivants: le contexte de la recherche, le thème de recherche, l'implication personnelle (motivation ou origine de l'intérêt pour le sujet), l'annonce des questions de départ et la problématique, le choix méthodologique et l'annonce du plan (Tran, 2010). Dans cet exemple, les MD, qui permettent de structurer ces informations, sont surlignés et la zone d'influence de chaque MD est délimitée entre crochets.

[...] Les nombreuses étapes que doivent franchir les jeunes apprenants francophones ou anglophones pour résoudre les difficultés grammaticales de l'orthographe de leur propre langue commencent à être connues (Thevenin, Totereau, Fayol, 1997 ; Nunes, Bryant et Bindman, 1997). **En revanche**, on ignore presque tout de ce processus quand il s'agit de l'apprentissage de la morphologie d'une langue étrangère enseignée dans le cadre scolaire.

**En guise d'introduction**, nous commencerons par présenter les principaux résultats concernant l'apprentissage du prétérit anglais par de jeunes Anglais, **puis** nous essayerons de cerner les conditions spécifiques caractérisant l'apprentissage de la morphologie verbale dans le cadre scolaire. **Ensuite**, nous proposerons une analyse contrastive des temps du passé en anglais et en français afin d'observer les secteurs de complexité potentielle pour les apprenants francophones. **Enfin**, nous présenterons une étude exploratoire de l'acquisition du prétérit anglais dans des textes écrits par des apprenants francophones. Nous la compléterons par une expérimentation construite à partir des hypothèses déduites de l'observation des formes verbales dans ces textes. Les résultats convergent pour suggérer que la conscience morphosyntaxique des apprenants **c'est-à-dire** leur capacité à prendre conscience des unités morphosyntaxiques et à les manipuler délibérément, se centre sur la forme des verbes français au détriment des conditions d'usage du prétérit en anglais.

Coralie Payre-Ficout, Jean-Pierre Chevrot, *La forme contre l'usage : étude exploratoire de l'acquisition du prétérit anglais*. Lidil 30, 2001, 101-116

**Figure 2 : Exemple de l'utilisation des MD dans l'introduction d'un article scientifique**

Dans cet extrait, l'auteur évoque le sujet et annonce le plan de l'article. L'enseignant peut laisser les apprenants relever les idées essentielles dans le texte et rassemble ensuite avec eux les idées. Il les aide à systématiser les points principaux à aborder dans l'introduction. À l'étape de production, l'enseignant peut faire rédiger une introduction sur papier ou à l'aide de ScribPlus<sup>7</sup> (Tran 2010), un outil TICE qui propose une aide logicielle pour la rédaction auprès

<sup>7</sup> Scribplus est un logiciel en ligne et accessible à l'adresse suivante : <http://scribplus.ovh.org/>. Il a été conçu par F. Mangelot, Lidilem, Université Grenoble Alpes et A. Khara, EIAL, Université Stendhal-Grenoble 3.

des étudiants étrangers. L'enseignant peut s'en inspirer pour concevoir des séquences didactiques qui correspondent aux objectifs et au niveau des apprenants. Le fait de repérer les structures discursives essentielles dans l'introduction apporte aux apprenants non seulement des connaissances sur la structuration de cette partie textuelle, mais les aide également à mieux observer d'autres textes. De plus, pour les activités de production écrite, ils peuvent réutiliser ces connaissances discursives et structurales dans leur écrit. Nous insistons sur le fait que le repérage des idées principales de chaque partie stratégique permet aux apprenants de mieux se concentrer sur leur écrit.

Le repérage des structures discursives sera facilité par les connaissances sur le fonctionnement syntaxique et sémantique des MD (les stratégies *bottom-up*). Dans l'exemple de la Figure 2, l'apprenant peut facilement repérer une succession d'idées développées grâce aux MD qui indiquent l'enchaînement thématique comme *puis, ensuite, enfin, etc.*, une reformulation introduite par *c'est-à-dire* ou bien une introduction de la nouvelle thématique comme *en guise de*. Ces deux stratégies *top-down* et *bottom-up* se complètent et seront bénéfiques pour améliorer les compétences rédactionnelles. Elles pourront participer à la construction des connaissances lexicales et structurelles d'une manière plus systématique.

Nous ne pouvons présenter ici que quelques grandes lignes sur l'introduction du corpus en classe de langue au moyen du corpus. Dans la pratique, il est important que l'enseignant utilise avec souplesse et en alternance différentes approches comme l'approche inductive, l'approche déductive, l'approche onomasiologique et l'approche sémasiologique.

## 6 Conclusion et perspectives

L'analyse des grammaires de FLE montre l'insuffisance de l'enseignement des MD dans le cadre des écrits universitaires. Cela nous a amenées à développer une réflexion sur ce sujet dans le but de comprendre leur fonctionnement linguistique dans les corpus et donc envisager leur enseignement pour un objectif précis qu'est la rédaction scientifique. Dans ce travail, nous avons effectué une analyse linguistique du marqueur de reformulation *c'est-à-dire*, en mettant en évidence son fonctionnement syntaxique et sémantique. Ces descriptions doivent passer par d'autres étapes de transposition didactique pour une adaptation adéquate aux étudiants non-natifs, par exemple, une simplification de la terminologie et une sélection des exemples adéquats pour chaque phénomène linguistique étudié. Comme notre intérêt porte sur le phénomène discursif, il est indispensable de faire travailler les apprenants sur les exemples plus larges pour qu'ils prennent conscience de l'organisation du texte grâce à ces MD. Dans la continuité de ce travail, nous envisageons de mettre en place des expérimentations auprès des locuteurs non-natifs du français afin de mieux adapter nos propositions didactiques au contexte d'enseignement.

## Bibliographie

- Anscombe, Jean-Claude (1983): « Pour autant, pourtant (et comment): à petites causes grands effets ». *Cahiers de linguistique française* 5: 37–84.
- Auchlin, Antoine (1981): « Marqueurs de structuration de la conversation et complétude ». *Cahiers de linguistique française* 1: 141–160.
- Boulton, Alex/Tyne, Henry (2014): *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*. Paris: Didier.

- Bourigault, Didier (2007): *Un analyseur syntaxique opérationnel : SYNTAX* (mémoire de HDR). Toulouse, Université de Toulouse.
- Burger, Harald (1998): *Phraseologie: eine Einführung am Beispiel des Deutschen*. Berlin: Schmidt.
- Cavalla, Cristelle/Loiseau, Mathieu (2013): « Scientext comme corpus pour l'enseignement ». In: Tutin, Agnès/Grossmann, Francis (eds.): *L'écrit scientifique : du lexique au discours. Autour de scientext*. Rennes, Presses Universitaires de Rennes: 163–180.
- Cavalla, Cristelle (2010): « Méthodologie d'apprentissage de l'écrit universitaire ». *Recherches et Applications* 47: 153–161.
- Un cadre européen commun de référence pour les langues (2001): *Apprendre, enseigner, évaluer*. Paris: Didier.
- Charolles, Michel (2003): « De la topicalité des adverbiaux détachés en tête de phrase ». *Travaux de linguistique* 47: 11–50.
- Charolles, Michel (1997): « L'encadrement du discours : univers, champs, domaines et espaces ». *Cahier de Recherche Linguistique* 6: 1–73.
- Charolles, Michel/Péry-Woodley, Marie-Paule (2005): *Les adverbiaux cadratifs*. Paris: Larousse.
- Chambers, Angela (2010): « L'apprentissage de l'écriture en langue seconde à l'aide d'un corpus spécialisé ». *Revue française de linguistique appliquée* 15: 9–20.
- Coxhead, Averil (2000): "A New Academic Word List". *TESOL Quarterly* 34/2: 213–238.
- Chetouani, Lamria (1988): *Vocabulaire général d'enseignement scientifique (VGES): recherche sur le vocabulaire de la communication scientifique orale en français dans les classes de lycée en Algérie* (Thèse de doctorat). Villetaneuse : Université de Paris XIII.
- Danjou-Flaux, Nelly (1980): « A propos du de fait, en fait, en effet, effectivement ». *Le français moderne* 48/1-2: 110–139.
- Gilquin, Gaëtanelle/Granger, Sylviane (2010): "How can data-driven learning be used in language teaching". *The Routledge handbook of corpus linguistics*: 359–370.
- Granger, Sylviane/Paquot, Magali (2008): "Disentangling the phraseological web". In: Granger, Sylviane/Meunier, Fanny (eds.): *Phraseology: An interdisciplinary perspective*. Amsterdam/Philadelphia, Benjamins: 27–49.
- Guimier, Claude (1996): *Les adverbes du français : le cas des adverbes en -ment*. Gap/Paris: Ophrys.
- Johns, Tim (1991): "Should you be persuaded: Two examples of data-driven learning". *English Language Research Journal* 4: 1–16.
- Landure, Corinne (2013): « Corpus et dictionnaires : prise de conscience linguistique chez des apprenants LANSAD ». *Apprentissage sur corpus : théories, méthodes, applications et perspectives* 97: 59–76.
- Mauranen, Anna (2004): "Speech corpora in the classroom". In: Aston, Guy/Bernardini, Silvia/Stewart, Dominic (eds.): *Corpora and Language Learners*. Amsterdam, Benjamins: 195–211.
- Muller, Claude (2005): « Portée et phrases négatives à modifieurs adverbiaux ». In: Choi-Jonin, Injoo/Bras, Myriam/Dagnac, Anne/Rouquier, Magali (eds.): *Questions de classification en linguistique : méthodes et descriptions*. Bern, Lang: 241–262.
- Muller, Claude (1991): *La négation en français*. Genève: Droz.

- Khatchatourian, Elisaveta (2008): « Les marqueurs de reformulation formés à partir du verbe dire ». In: Le Bot, Marie-Claude/Shuwer, Martine/Richard, Elisabeth (eds.): *La reformulation : marqueurs linguistiques, stratégies énonciatives*. Rennes, Presses universitaires de Rennes: 19–33.
- Paquot, Magali (2010): *Academic vocabulary in learner writing: from extraction to analysis. Research in corpus and discourse*. London/New York: Bloomsbury.
- Paillard, Denis/Vu, Thi Ngan (2012): *Inventaire raisonné des marqueurs discursifs du français. Description. Comparaison. Didactique*. Paris: AUF.
- Pecman, Mojca (2004): *Phraséologie contrastive anglais-français: analyse et traitement en vue de l'aide à la rédaction scientifique*. (Thèse de doctorat). Nice: Université Sophia Antipolis.
- Phal, André (1971): *Vocabulaire général d'orientation scientifique*. Paris: Crédif.
- Rossari, Corinne/Jayez, Jacques (1997): « Connecteurs de conséquence et portée sémantique ». *Cahiers de linguistique française* 19: 233–266.
- Rossari, Corinne (1994): « Les opérations de reformulation: analyse du processus et des marques dans une perspective contrastive français-italien ». *Sciences pour la communication* 40: x–225.
- Rossari, Corinne (1990): « Projet pour une typologie des opérations de reformulation ». *Cahiers de linguistique française, Marquage linguistique, inférence et interprétation dans le discours* 11: 345–359.
- Roulet, Eddy (1987): « Complétude interactive et connecteurs reformulatifs ». *Cahiers de linguistique française* 8: 111–140.
- Steuckardt, Agnès/Niklas-Salminen, Aïno (eds.) (2005): « Les marqueurs de glose ». *Langues et langage* 12/1. Aix-en-Provence, Publications de l'Université de Provence.
- Tran, Thi Thu Hoai (2014). *Développement d'une aide à l'écrit scientifique. Description de la phraséologie scientifique et réflexion didactique pour l'enseignement à des étudiants non natifs*. (Thèse de doctorat). Grenoble: Université Grenoble Alpes.
- Tran, Thi Thu Hoai (2010): *Aides logicielles à la rédaction universitaire. L'exemple des collocations*. Mémoire Master 2 Recherche FLE. Grenoble, Université Stendhal Grenoble 3.
- Tutin, Agnès (2014): « La phraséologie transdisciplinaire des écrits scientifiques : des collocations aux routines sémantico-rhétoriques ». In: Tutin, Agnès/Grossmann, Francis (eds.): *L'écrit scientifique : du lexique au discours. Autour de scientext*. Rennes, Presses Universitaires de Rennes: 27–44.
- Tutin, Agnès (2008): « Sémantique lexicale et corpus : l'étude du lexique transdisciplinaire des écrits scientifiques ». *Lublin Studies in Modern Languages and Literature* 32: 242–260.
- Tutin, Agnès (2007): « Traitement sémantique par analyse distributionnelle des noms transdisciplinaires des écrits scientifiques ». *Actes de Traitement Automatique des Langues Naturelles (TALN)*: 283–292.
- Vassiliadou, Hélène (2008): « Quand les voies de la reformulation se croisent pour mieux se séparer : à savoir, autrement dit, c'est-à-dire, en d'autres termes ». In: Le Bot, Marie-Claude/Shuwer, Martine/Richard, Elisabeth (eds.): *La reformulation : marqueurs linguistiques, stratégies énonciatives*. Rennes, Presses universitaires de Rennes: 35–49.