

Guides touristiques et combinatoire verbale

Resumé

Cet article propose de réfléchir sur la question de la combinatoire verbale présente dans le discours touristique, en analysant quelques exemples de collocations tirées d'un corpus de guides touristiques en langue française et comparées à leurs équivalents en langue italienne et portugaise, appartenant aussi à un corpus du même type. Il s'agit des premiers pas d'une étude qui se veut comparative trilingue et qui permettra d'enrichir la base de données lexicographiques du projet de recherche LBC – *Lessico dei Beni Culturali* de l'Université de Florence.

Mots clés : guide touristique, combinatoire verbale, collocation

Resumo

Este artigo propõe uma reflexão sobre a questão da combinatória verbal presente no discurso turístico. Analisar-se-á alguns exemplos de colocações provenientes de um corpus que reúne guias turísticos em língua francesa e que serão confrontados com os seus equivalentes em língua italiana e portuguesa, também provenientes de um corpus do mesmo tipo. Trata-se dos primeiros passos de um estudo que se quer comparativo trilingue e que permitirá enriquecer a base de dados lexicográficos do projecto de investigação LBC – *Lessico dei Beni Culturali* – Università degli Studi di Firenze.

Palavras chave : guia turístico, combinatória verbal, colocação

* CLUNL-FCSH, Universidade NOVA de Lisboa

1. Introduction

À l'instar d'autres types et genres textuels, les guides touristiques fourmillent de clichés langagiers propres à ce genre d'écrits, se manifestant, entre autres, dans la combinatoire des différentes unités lexicales/terminologiques de ce discours.

Sur la base de ce constat, nous nous proposons, dans cet article, de présenter la méthodologie de travail ainsi que les premiers résultats d'une étude¹ de la combinatoire verbale du discours touristique.

Notre approche se penche en particulier sur la collocation verbale V+N, en sachant que, dans la définition que nous adoptons, la collocation est une combinaison non libre constituée d'une base (N = terme) choisie librement et d'un collocatif (ici, verbe) qui permet d'attribuer un sens spécifique à l'expression et/ou qui l'encadre dans un discours plus ou moins spécialisé (Dechamps 2013a).

Par ailleurs, cette approche s'appuie sur l'analyse de données collocationnelles issues d'un corpus électronique trilingue (français – italien – portugais), constitué expressément pour cette recherche et réunissant différents guides touristiques collectés sur Internet.

Notre objectif est de démontrer 1) l'importance de l'étude des phénomènes collocationnels en langue de spécialité et 2) la nécessité de l'intégration de cette information dans les dictionnaires ou base de données terminologiques afin de permettre une meilleure rédaction et/ou traduction des textes touristiques.

Ainsi, dans cet article, après une brève exposition de la problématique en analyse, nous décrivons notre projet de recherche, le corpus textuel qui a été créé à cet effet ainsi que les premiers résultats obtenus concernant principalement la langue française. En guise de conclusion, nous énoncerons quelques perspectives de recherche sur la base de ces mêmes résultats.

2. Problématique

L'importance de la traduction spécialisée est depuis longtemps reconnue et il serait erroné de penser que celle-ci se limite à trouver des équivalents aux termes contenus dans le texte. En fait, il est essentiel de considérer des phénomènes qui relèvent de la combinatoire, appelés par certains, syntagmes, phraséologies, phrasèmes ou collocations. Cette dernière dénomination reprise à des linguistes comme Marie-Claude L'Homme (2004 : 112-114) sera celle adoptée dans cet article. Ce type de structures représente en réalité un défi pour le traducteur, quelle que soit la langue de spécialité qui est en jeu.

¹ Étude financée par la FCT – Fundação para a Ciência e Tecnologia, dans le cadre du projet UID/LIN/03213/2013 du Centro de Linguística da Universidade NOVA de Lisboa.

Dans le prolongement d'autres travaux déjà réalisés dans le contexte de la langue juridique (FR-PT), nous avons pris le parti de nous attarder sur la combinatoire verbale présente dans le discours touristique, consciente de sa pertinence dans le cadre d'une réflexion plus générale sur l'emploi des verbes en langue de spécialité.

Dans ce sens, les approches proposées par Marie-Claude L'Homme (2012, 2015) et Gaston Gross (2010) nous paraissent particulièrement utiles pour comprendre le fonctionnement des verbes dans les discours spécialisés. Les structures actancielles (L'Homme), inspirées de la Sémantique des cadres², comme les schémas d'arguments (Gross), permettent de mieux mettre en évidence les régularités au sein des choix terminologiques à opérer au moment de combiner entre elles les différentes unités lexicales (UL) et/ou terminologiques (UT) afin d'attribuer un sens spécifique à la collocation. Ou, suivant la perspective de Gross (2010 : 188), « c'est le contexte qui détermine le sens d'un terme ». D'une façon ou d'une autre, dans cette optique, on élargit le concept d'unité de sens qui ne se limite plus ainsi à l'UL/UL, même complexe, mais à un ensemble d'unités qui entretiennent des relations privilégiées dans le cadre de la phrase simple.

Concrètement et en observant des exemples relativement simples, il faut noter que :

- Certains verbes acquièrent, dans le discours spécialisé, une acception spécialisée en contact avec la base de la collocation.

Prendre = saisir
Prendre une photo = photographier

- Il existe une certaine restriction au niveau de la sélection du collocatif. Des verbes apparemment synonymes ne sont pas facilement interchangeables et le choix de ceux-ci dépendront beaucoup du terme, base de la collocation.

photographier
prendre une photo(graphie)
faire une photo(graphie)
* tirer une photo(graphie) vs tirar uma fotografia (PT) / scattare una fotografia (IT)

photocopier
faire une photocopie

² Frame Semantics en anglais. Voir Fillmore 1982.

- * prendre une photocopie
- ? tirer une photocopie

De plus, cette restriction, par le choix de collocatifs spécifiques, inscrit la combinaison dans un discours plus ou moins spécialisé.

- Contracter une assurance de voyage (discours spécialisé)
- Souscrire une assurance de voyage
- Prendre une assurance de voyage (discours de vulgarisation)

Toutes ces observations démontrent l'importance d'un plus grand investissement dans une description terminographique des collocations plus systématique et complète dans les ouvrages de référence consultés par les rédacteurs et les traducteurs.

3. Le projet de recherche

Notre étude sur la combinatoire verbale du discours touristique s'inscrit dans les axes de recherche du projet LBC – *Lessico dei Beni Culturali* – centralisé à l'Université de Florence (Italie) et s'organise en trois phases. La première concerne l'élaboration d'un corpus textuel trilingue comparable, regroupant des guides touristiques en français, en portugais et en italien. La deuxième s'occupe de la description des principales collocations relevées dans le corpus, et la troisième de l'enrichissement de la base de données terminologiques LBC³ en y complétant l'information collocationnelle sur la base de documents authentiques et contemporains.

3.1. Le corpus LBC – FR/PT/IT

Pour satisfaire les objectifs de cette recherche sur le discours touristique, un corpus comparable trilingue qui réunit des guides de voyage collectés sur Internet a été élaboré. Concrètement, il s'agit de textes relativement brefs décrivant les principales attractions d'une ville – dans ce cas, Paris, Lisbonne ou Florence – et mis à la disposition des voyageurs sur différents sites consacrés au tourisme.

Ces textes s'inscrivent tous dans le discours de vulgarisation où l'on retrouve principalement des séquences descriptives et explicatives (Adam 2011), avec une simplification de structures syntaxiques et une certaine sélection terminologique. On observe également d'autres marques de didacticité, en reprenant les termes de Beacco (1995) et Moirand (1993), marques

³ Une ébauche de cette base de données terminologiques est consultable à l'adresse suivante : <<http://www.lessicobeniculturali.net/fr>> (03/16).

qui ne se limitent pas au seul discours scientifique pédagogique ou didactique, mais qui sont observables dans tout discours où il y a transmission de connaissances, comme c'est le cas du discours qui caractérise les guides touristiques. Globalement, il s'agit d'un ensemble de reformulations, aussi bien au niveau iconique (plans, tableaux, illustrations, photos) que verbal.

Ce corpus est divisé en trois parties thématiques : Florence, Lisbonne et Paris, mais aussi en trois aires linguistiques : français, italien et portugais. Actuellement ce corpus comprend pour chacune de ses parties le nombre de mots suivants :

Corpus LBC	Français	Italien	Portugais
Paris	75 719 mots	25 347 mots	18 021 mots
Lisbonne	21 822 mots	6 286 mots	2 856 mots
Florence	27 698 mots	15 701 mots	---
TOTAL	125 239 mots	47 334 mots	20 877 mots

Les sous-corpus portugais et italien sont en ce moment encore très réduits, mais nous pensons les développer à plus ou moins court terme. Il faut souligner en effet la difficulté à trouver sur Internet des textes en portugais européen appartenant à la catégorie choisie. Les textes recueillis jusqu'à aujourd'hui appartiennent tous à la variante brésilienne.⁴

Le logiciel dont on s'est servi pour le traitement des données textuelles est Hyperbase 9.0 (2010) conçu par Etienne Brunet de l'Université de Nice. Ce logiciel a l'avantage d'offrir un traitement spécifique pour chaque langue étudiée ici, en permettant notamment la confrontation des données recueillies à un corpus de référence. En ce qui concerne la lemmatisation, celle-ci est prise en charge par le programme TreeTagger.

Pour compléter le traitement fourni par Hyperbase, le logiciel TermoStat 3.0 créé par Patrick Drouin de l'Université de Montréal⁵, a également été utilisé, notamment pour déterminer les combinaisons N+V / V+N (« bigrammes ») les plus spécifiques attestées dans ce corpus.

Les points forts de ce corpus sont, d'une part, son homogénéité et, d'autre part, sa richesse textuelle et terminologique. On est assurément

⁴ Soulignons que notre objectif est de constituer un corpus qui soit représentatif de la diversité de chaque aire linguistique, en ne privilégiant aucune norme ou variante en particulier. Avec Hyperbase, il est techniquement possible d'identifier l'origine de chaque texte (et, de là, les extraits et les occurrences fournis par le logiciel lors de l'analyse). Cette information sur la norme ou variante est tout à fait pertinente pour enrichir la description terminologique et collocationnelle.

⁵ Disponible à l'adresse suivante : < <http://termostat.ling.umontreal.ca/> > (03/16)

face à un ensemble de textes appartenant au même genre textuel, avec le même degré de spécialisation et rédigés lors de la même période temporelle (2010-2015). Ce corpus, de par ses caractéristiques, se présente comme une source précieuse de contextes authentiques et contemporains pour la description terminographique, notamment pour l'insertion d'exemples d'emploi de termes et de collocations terminologiques dans la base de données LBC. Il est toutefois loin d'être parfait et comporte quelques fragilités à corriger. La première vulnérabilité à relever est la taille réduite du corpus. Dans quelle mesure est-il réellement représentatif de la langue utilisée dans ce contexte ? La deuxième est en rapport avec le déséquilibre entre les différentes langues concernées dans cette étude ; ce qui soulève un certain nombre de questionnements dans l'optique d'une étude comparative.

4. Description des collocations : premiers résultats

Malgré les diverses limitations du corpus évoquées ci-dessus, quelques résultats dignes d'intérêt concernant spécifiquement la langue française ont été obtenus, ouvrant la voie vers des recherches plus ambitieuses qui renforceraient, entre autres, l'étude comparative des phénomènes collocationnels dans les trois langues.

En termes méthodologiques, les démarches suivantes ont été adoptées. Dans un premier temps, nous sommes partie du terme « musée » – terme le plus fréquent après « rue » et « place » mais le plus spécifique des trois – dans le sous-corpus français.⁶ D'autres termes se trouvant dans ce même corpus et appartenant au même domaine conceptuel ont été sélectionnés et nous leur avons attribué une étiquette sémantique qui a rendu possible leur répartition en classes d'objets (Gross).

Classes d'objets	Termes
<espace muséal>	musée, galerie
<ensemble de biens culturels>	collection
<présentation de biens culturels>	exposition
<bien culturel>	œuvre, chef d'œuvre (hyperonymes) tableau, peinture, fresque, sculpture, statue (hyponymes)

Cette approche nous a permis d'élaborer une première ébauche de ce domaine conceptuel, complétée par l'étude de Mariannick Jadé (2011) sur la terminologie du patrimoine.

⁶ Le nombre d'occurrences pour chacun des termes est le suivant : rue (417), place (325) et musée (278). Le score de spécificité est 66,5, 38,66 et 79,3 respectivement. Source : TermoStat 3.0.

Dans un deuxième temps, sur la base de ces différentes unités terminologiques, nous avons amorcé l'analyse de la combinatoire verbale, en privilégiant les structures transitives directes. Cette option nous a conduite à étudier les structures à la voix active et passive mais aussi l'emploi adjectival du participe passé.

Souscrire une assurance de voyage⁷

Le touriste souscrit une assurance de voyage

Une assurance de voyage a été souscrite par le touriste

L'assurance souscrite ...

Il faut souligner que les fonctions « bigrammes » du programme TermosStat Web 3.0 et « concordance » / « contexte » du programme Hyperbase 9.0 ont été particulièrement utiles. La fonction « cooccurrence » d'Hyperbase n'a pas été sollicitée dans le cadre de cette étude, vu qu'elle peut être efficace pour le relevé des unités terminologiques complexes mais pas vraiment pour celui des collocations.

En réalité, décrire la combinatoire verbale est un processus difficile, même avec l'aide de programmes informatiques de traitement de données textuelles. L'analyse manuelle reste encore nécessaire. Les principales pierres d'achoppement sont en effet :

1) la distance entre la base et le collocatif par la présence de compléments et de modificateurs

La basilique de Santa Croce abrite en fait un nombre incroyable de chefs d'oeuvre. (Corpus LBC FR - FL)

À ce propos, Gross (2010 : 189) affirme qu'« il y a souvent entre un prédicat et ses arguments des insertions de diverses natures qui n'appartiennent pas à la structure pertinente et qu'il faut ignorer si l'on veut retrouver la distribution qui éclaire le terme en question ».

2) les anaphores

Le musée du Quai Branly, que l'on appelle aussi souvent musée des Arts Premiers est en fait le musée des arts et civilisations d'Afrique, d'Asie, d'Océanie et des Amériques. Il est abrité dans une immense réalisation due à Jean Nouvel qui mêle technologie et nature [...] (Corpus LBC FR - PA)

⁷ « Si vous avez souscrit une assurance contre le vol de vos papiers, déclarez-les avec exactitude. » (Corpus LBC FR - PA)

D'ailleurs, à ce propos, Marie-Claude L'Homme (2004 : 80) affirme que :

L'anaphore est un phénomène complexe et particulièrement difficile à modéliser dans un traitement automatique en raison de son imprévisibilité. Cette imprévisibilité intervient à deux niveaux puisqu'on ne peut pas anticiper avec certitude ni la forme qu'elle peut prendre, ni l'endroit dans le texte où elle peut apparaître.

De plus, déterminer le degré de cohésion entre les deux éléments de la collocation est une tâche complexe. Toutefois, le logiciel TermoStat 3.0, sur la base de statistiques lexicales et de la confrontation de notre corpus spécialisé à un corpus de référence non spécialisé, propose la liste suivante des « bigrammes » les plus spécifiques pour le domaine conceptuel étudié :

Bigrammes	Fréquence	Score d'association
Visiter un musée	7	21.50
Abriter une collection	8	21.18
Accueillir une exposition	4	14.46

En ce qui concerne « abriter », verbe de l'un des trois bigrammes fournis par TermoStat 3.0, deux aires sémantiques principales peuvent être déterminées : celle de l' /accueil/ où « abriter » peut être équivalent à « accueillir », « recevoir » et « héberger » et celle de la /protection/ où « protéger », « conserver » et « préserver » peuvent être considérés comme des (para) synonymes de ce même verbe « abriter ». Comme la plupart des verbes, « abriter » est polysémique (Gross 2010 : 188).

ABRITER	/accueil/	Accueillir Recevoir Héberger
	/protection/	Protéger Conserver Préserver

Ainsi, en reprenant le corpus LBC, nous retrouvons des séquences telles que « un musée abrite/accueille/conservé une collection » mais aucune des combinaisons suivantes n'est attestée : « un musée reçoit/héberge/préserve/protège une collection ». Avec cet exemple, on observe facilement les préférences collocationnelles opérées dans ce type d'écrits et la difficulté à substituer un verbe par un autre (para)synonyme, difficulté qui est due à des contraintes propres au discours.

À partir de la consultation des occurrences du verbe « abriter » dans le corpus, nous pouvons également déterminer les schémas d'arguments suivants :

Schémas d'arguments	Exemples retirés du corpus LBC - FR
<espace 'fermé'> abrite <espace 'fermé'>	Un édifice abrite une station de métro, une gare
<espace 'fermé'> abrite <espace muséal>	La Galleria della Accademia/le palais du Louvre/le palais de Tokyo abrite un musée
<espace 'fermé'> abrite <institution>	Un édifice abrite le podestat/la compagnie d'Orsanmichele/la Bourse
<espace 'fermé'> abrite <humain>	Un café, un studio abrite des artistes
<espace 'fermé'> abrite <bien culturel><un ensemble de biens culturels><présentation de biens culturels>	Une loge abrite une œuvre La Galleria Palatina /la Cappella degli Ubriachi /la Galleria del Costume abrite une collection/une œuvre/un chef d'œuvre
<espace muséal> abrite <œuvre><un ensemble de biens culturels><présentation de biens culturels>	Un musée abrite une collection, une exposition, une œuvre
<espace 'ouvert'> abrite <commerce><institution>	Un quartier/une avenue/une rue abrite des bars/ des restaurants/des discothèques/des librairies/un ministère
<espace 'ouvert'> abrite <humain><profession>	Un quartier abrite des habitants Le Ponte Vecchio abrite des orfèvres, des joailliers

En résumant l'information que comporte ce tableau et en appliquant le même type de recherche sur les autres verbes (para)synonymes – « accueillir »/« recevoir »/« héberger » et « conserver »/« préserver » –, on peut arriver à la synthèse qui suit et qui met en perspective la combinatoire des verbes sélectionnés.

sujet	Verbe	COD
<espace fermé> <espace muséal>	abriter	<espace muséal> <présentation de biens culturels> <ensemble de biens culturels>
	accueillir	<bien culturel>
	conserver	<ensemble de biens culturels> <bien culturel>
	recevoir	<humain> <touristes>
<espace ouvert>	préserver	<bien naturel> <bien culturel>
<espace fermé>	héberger	<humain> <touristes>

Dans une approche trilingue et en se concentrant plus particulièrement sur les verbes « abriter » et « accueillir », les plus fréquents des six verbes

présentés, nous remarquons les équivalents italiens et portugais sont « ospitare/accogliere » et « abrigar/acolher », le premier verbe de chaque série comportant un trait sémantique lié au caractère permanent de l' 'accueil' et, de là, à l'idée de protection, tandis que, pour le second, ce trait renvoie au caractère souvent non permanent de ce même « accueil », surtout en se référant, par exemple, à des biens culturels présentés lors d'une exposition temporaire.

Le musée abrite également une exposition permanente et a une section consacrée aux non-voyants. (Corpus LBC FR – FL)

Il Museo Nazionale di Archeologia ospita la più importante collezione archeologica del Portogallo. (Corpus LBC IT – LI)

O Museu do Louvre abriga um acervo valiosíssimo de obras de arte de vários géneros e épocas. (Corpus LBC PT – PA)

La Pinacothèque de Paris accueille des expositions temporaires de niveau international. (Corpus LBC FR – PA)

La Cité de la Musique accoglie un museo, mostre temporanee e una grande sala di spettacolo. (Corpus LBC IT – PA)

Visite o Castelo d'Auvers, mansão do século XVII que acolhe exposições sobre os impressionistas. (Corpus LBC PT – PA)

Si, selon Gross (2010 : 192), « c'est la nature des arguments qui permet de discriminer les différents polysèmes »⁸, c'est aussi la nature de ces mêmes arguments – et, plus précisément leurs caractéristiques sémantiques –, qui déterminera le verbe à choisir parmi plusieurs (para)synonymes.

5. Conclusions préliminaires et perspectives de recherche

La brève étude exposée dans cet article mérite plusieurs observations.

Le corpus utilisé comporte le désavantage d'être de taille limitée et les premiers résultats obtenus sont évidemment à considérer avec prudence.

Toutefois nous insistons sur le fait que ce corpus est loin d'être statique et qu'il est prêt à recevoir d'autres textes pour renforcer la représentativité de la langue de spécialité analysée et pour corriger éventuellement les résultats déjà obtenus. Cet enrichissement textuel pourra se concrétiser avec l'intégration 1) de textes du même type, c'est-à-dire des guides touristiques, 2) d'autres types de textes qui présentent différents degrés de spécialisation et 3) des textes provenant d'autres pays francophones et lusophones.

Par ailleurs, cette brève analyse autour du terme « musée » et du verbe « abriter » est un premier pas vers des études plus ambitieuses et démontre parfaitement la pertinence de l'étude des collocations verbales dans l'op-

⁸ Comme dans *prendre un billet de train, prendre un avion, prendre une assurance de voyage, prendre une photo* où le verbe aura un sens différent suivant l'argument COD qui le suit.

tique de développement de bases de données lexicographiques et/ou terminographiques, en offrant au traducteur et au rédacteur des informations utiles souvent absentes des ouvrages lexicographiques/terminographiques couramment consultés.

6. Références bibliographiques

- Adam J.-M. 2011, *Les textes : types et prototypes*, 3rd ed., Armand Colin, Paris.
- Beacco J.-C., Moirand S. 1995, « Autour des discours de transmission des connaissances », *Langages*, 117 : 32-53.
- Dechamps C. 2013a, *Les collocations de la langue juridique française: problématiques de l'enseignement/apprentissage à des apprenants lusophones*, thèse de doctorat non publiée, Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas.
- Dechamps C. 2013b, « L'enseignement/apprentissage du français juridique : une proposition de démarche terminologique et didactique », *Équivalences*, 40/1-2 : 191-224.
- Fillmore C. 1982, *Frame Semantics*, in *Linguistics in the Morning Calm*, Hanshin Publishing, Seoul : 111-137.
- Gross G. 2010, « Sur la notion de contexte », *Meta*, 55-1 : 187-197.
- Jadé M. 2011, « Contribution aux réflexions sur les terminologies du patrimoine : méthodologie et proposition. L'exemple du patrimoine scientifique et technique » <https://f.hypotheses.org/wpcontent/blogs.dir/1495/files/2013/08/terminologie_pat_jade.pdf> (03/16).
- L'Homme M.-C. 2004, *La terminologie : principes et techniques*, Presses Universitaires de Montréal, Montréal.
- L'Homme M.-C. 2012, « Le verbe terminologique : un portrait de travaux récents, in Actes du Congrès Mondial de Linguistique », EDP Sciences : 93-107 <http://www.shs-conferences.org/articles/shsconf/pdf/2012/01/shsconf_cmlf12_000340.pdf> (03/16).
- L'Homme M.-C. 2015, « Découverte de cadres sémantiques dans le domaine de l'environnement : le cas de l'influence objective », *Terminalia* 12 <<http://www.raco.cat/index.php/Terminalia/article/view/305114/394949>> (03/16).
- Moirand S. 1993, « Autour de la notion de didacticité », *Les Carnets du Cediscor* 1 : 9-20.