

Contexte et sens: utilisation d'un corpus écrit dans l'enseignement / apprentissage du FLE

Alain KAMBER

Institut de langue et civilisation françaises, Université de Neuchâtel

The aim of the present paper is to show that the use of a very large electronic corpus is likely to simplify the foreign language teacher's task as well as the learner's. Using the 'Corpus français', we present two examples of tasks which can be achieved by students of French as a foreign language even though they do not necessarily have in-depth knowledge of linguistic analysis and for which no prior filtering by the teacher is needed. In the first task, the aim is to determine the typical use of a verb by looking for co-occurrences strongly collocating with it, while in the second one, the purpose is to disambiguate near-synonyms, based on the usage contexts specific to each of the lexical units considered. It appears that the results obtained compare favourably with the descriptions found in dictionaries, despite the simplicity of the procedure.

1. Introduction

Si de nombreux auteurs s'accordent à dire qu'il faudrait ménager une place plus importante aux données et aux corpus dans l'enseignement des langues étrangères, respectant ainsi le *data-driven learning* que Johns (1991) appelle de ses vœux, force est de constater que la pratique est encore souvent fort éloignée de ce postulat d'apprentissage / enseignement à partir de données brutes, et de "l'exploitation directe des corpus par l'apprenant" (Boulton, 2008). En effet, si la plupart des enseignants de langue étrangère adhèrent au postulat d'authenticité¹ et s'accordent sur la nécessité de confronter l'apprenant à une langue naturelle, l'intuition se révélant sans valeur scientifique, souvent, la pratique de ces enseignants consiste à compiler des résultats de recherches qu'ils ont eux-mêmes effectuées, pour ensuite livrer ces résultats à leurs étudiants sous forme de concordances, c'est-à-dire de listes comportant les occurrences d'une unité-pôle entourée de ses contextes d'utilisation. Si une telle approche didactique peut sans doute s'avérer efficace – encore qu'elle nous semble exiger des compétences d'analyse de linguiste dont ne disposent pas nécessairement les apprenants – elle ne semble en revanche guère contribuer au renforcement de l'autonomie de l'apprenant, puisque l'enseignant, même s'il renonce à

¹ On pensera ici à Sinclair (1991: 6): "One does not study all of botany by making artificial flowers". Si ce principe est valable pour l'étude linguistique, il ne peut que l'être à plus forte raison pour l'apprentissage d'une langue.

faire office de "guide" et de "mentor" (Willis, 2003: 167), continue de s'attribuer un rôle d'intermédiaire indispensable entre le corpus – c'est-à-dire les faits de langue (Saussure, 1986) – et l'apprenant. Quant à la pratique consistant à charger les étudiants de faire eux-mêmes ce type de recherche sur des sites livrant une représentation en KWIC (*keyword in context*), en les chargeant de dégager ensuite des règles de nature lexicale ou syntaxique (cf. Duda & Tyne, 2010: 95ss.), si elle nous semble largement préférable pour renforcer l'autonomie de l'apprenant, elle présente le risque de rétrécir son champ de vision, puisqu'il ne s'intéressera plus qu'aux mots se trouvant à proximité immédiate du mot-clé, et non à son contexte d'utilisation général.

Dans la présente contribution, nous nous proposons de montrer comment l'utilisation d'un corpus "grand public", avec des outils de recherche extrêmement simples à manier, est susceptible de permettre aux apprenants du français un accès direct à la réalité langagière écrite, et à une interprétation approfondie des données sans disposer obligatoirement d'un savoir étendu de linguiste. Deux exemples touchant au domaine du lexique-grammaire nous permettront d'illustrer cette pratique: le premier concerne l'usage du verbe *entourer* et sa complémentation, le second les contextes d'utilisation de quelques quasi-synonymes de *regarder* et la désambiguïsation de leur sens.

2. Un corpus simple à utiliser

Contrairement à d'autres langues, pour lesquelles il existe un certain nombre de corpus divers et de grande dimension², il n'existe, pour le français, que peu de grandes banques de données "accessibles de façon systématique" (Leech, 1991). Ainsi, *Frantext* est certes un outil précieux et de dimensions appréciables avec ses 210 millions de mots, mais il présente un certain nombre de limites: cet instrument est constitué en grande partie de références littéraires, et son accès est limité par un système d'abonnement. En outre, il présente une dimension diachronique qui s'accorde mal avec les besoins d'apprenants du FLE³. Le souhait de disposer d'outils supplémentaires et complémentaires est donc régulièrement exprimé, aussi bien par les linguistes que par les enseignants.

² On pensera notamment à l'anglais ou à l'allemand, pour lesquels des corpus de très grande taille existent depuis plusieurs années et sont accessibles sans restrictions à toute personne intéressée, par exemple DeReKo – Das Deutsche Referenzkorpus – qui compte environ 4.1 milliards de mots pour l'allemand (état au 29.03.2011).

³ Dans sa présentation générale de la linguistique de corpus, Williams (2006: 152) rappelle à juste titre que tout corpus a été créé "afin de résoudre des problématiques différentes."

Le corpus utilisé dans cette recherche est *Corpus français*⁴ (désormais *CF*), une banque de données de français contemporain comprenant 700 millions de mots mise en ligne en janvier 2010⁵. Le défi relevé par ce corpus est double: tout d'abord, mettre à disposition "une collection de morceaux de texte en langues sous format électronique, sélectionnés en fonction de critères externes afin de représenter, dans la mesure du possible, une langue ou variété de langue comme source de données pour la recherche linguistique" (Sinclair, 2005, traduit par Williams, 2008). Ensuite, donner à cette collection un volume suffisant pour pouvoir en tirer des résultats représentatifs selon le principe "More data is better data" (Church & Mercer, 1993). Sans constituer en aucun cas un corpus "équilibré" (Péry-Woodley, 1995), *CF* comprend ainsi des textes de différentes provenances et surtout dispose d'une masse suffisante pour être considéré comme représentatif de la langue écrite contemporaine.

CF est constitué de trois sous-corpus: corpus journalistique, corpus Internet et corpus Wikipédia. On peut consulter chacun de ces corpus séparément ou les trois ensemble. Imaginé pour différents publics – chercheurs comme utilisateurs de la langue à la recherche d'informations sur l'emploi du lexique en contexte –, il a été conçu de manière à fournir pour chaque mot-clé recherché un certain nombre d'informations. À côté d'indications relatives à sa fréquence (nombre d'occurrences, classe de fréquence) apparaissent d'abord de nombreuses phrases exemples, avec les références des sources, puis une série d'informations sur l'utilisation du terme en contexte: plusieurs listes de mots mettant en lumière ses affinités syntagmatiques sont automatiquement calculées et affichées (log-likelihood): cooccurrences significatives au sein d'une phrase, voisins significatifs de gauche et de droite. Une représentation graphique visualise les liens de proximité et de fréquence.

Les outils de consultation proposés par *CF* présentent un certain nombre de limites. En effet, cette banque de données ne permet des recherches que sur le plan de l'unité lexicale, et les résultats obtenus restent tous restreints au niveau de la phrase graphique, sans possibilité systématique d'accéder au cotexte étendu. L'absence de lemmatiseur, l'impossibilité d'exporter des données et de combiner ou d'affiner les critères de recherche constituent des obstacles supplémentaires. Il n'en demeure pas moins que, malgré tous ces désavantages, l'utilisation de ce corpus dans l'enseignement / apprentissage du FLE reste possible et peut donner des résultats tout à fait intéressants⁶.

⁴ http://wortschatz.uni-leipzig.de/ws_fra/

⁵ Sur le projet Wortschatz-Portal, voir Quasthoff, Richter & Biemann (2006).

⁶ Sur le lien entre corpus et types d'utilisation possibles, voir Charaudeau (2009).

3. Applications didactiques concrètes

On sait que les verbes imposent des restrictions syntaxiques. Ainsi, parmi plusieurs constructions possibles, le verbe *rendre* en présente une dans laquelle il régit un double complément, direct et indirect: *Marie vous rendra sûrement votre livre demain* (Tesnière, 1959: 127). La connaissance et la maîtrise d'une telle règle permettent certes théoriquement à l'apprenant de produire des énoncés grammaticalement corrects, en utilisant notamment la bonne préposition pour le complément indirect. Mais, au-delà de ces restrictions syntaxiques, les verbes connaissent également des restrictions sémantiques, que l'on peut classer en catégories, comme cela est pratiqué notamment par le laboratoire LADL de l'Université de Paris VII⁷. On définira ainsi les catégories suivantes:

[hum]:	"humain"	>	la dame, la fille, notre maîtresse, etc.
[coll]:	"collectif"	>	le gouvernement, le comité, la commission, etc.
[anim]:	"animé"	>	l'homme, le chien, le ver de terre, etc.
[concr]:	"concret"	>	le livre, la voiture, un téléphone, etc.
[abstr]:	"abstrait"	>	une opinion, la volonté, un résultat, etc.
[lieu]:	"lieu"	>	à la maison, de Rome, etc.

Fig. 1: Différentes catégories sémantiques

Dans un contexte d'enseignement / apprentissage de FLE, ces catégories permettront d'expliquer par exemple que si des énoncés comme *L'homme court* ou *La gazelle court* sont corrects, en revanche **La cathédrale court* n'est pas acceptable. En résumé, si le verbe *courir* accepte des sujets des types [hum] et [anim], en revanche il ne tolère pas un sujet [concr]. Il s'agit certes là d'apports importants, mais ce type d'explications ne permet pas de toucher à l'essentiel. En effet, on peut sans conteste approfondir l'analyse et parvenir à des résultats plus satisfaisants en appliquant le principe de Firth (1957: 11): "... you shall know a word from the company it keeps". Les mots considérés individuellement – et non simplement des catégories de mots – présentent en effet des affinités entre eux. Il s'agira donc de déterminer quels sont les partenaires privilégiés d'une unité lexicale en partant du principe de l'hypothèse minimale (Sinclair, 1986 et 1991) et en accordant au corpus le rôle central dans une recherche inductive (Tognini-Bonelli, 2001)⁸. Dans cette perspective, on accordera certes une attention soutenue au "choix lexical des noms admis en position de sujet ou de complément", (Sánchez Cárdenas, 2010: 42), mais on tiendra également compte des autres éléments gravitant autour du verbe, et qui

⁷ Voir notamment Gross (1975), ainsi que Harris (1976). Voir également Picoche (1993: 72ss.) et Blanche-Benvéniste (2002).

⁸ Voir également l'article de Steyer dans ce numéro.

contribuent eux aussi à son sens, tels les adverbiaux, les compléments (de but, de moyen, etc.) et les semi-auxiliaires (de causalité, d'aspect, de modalité, etc.). Pour pouvoir observer ces derniers, on introduira donc la forme infinitive des verbes comme terme de recherche.

3.1 Contexte d'utilisation, sens d'un mot: entourer

L'apprenant, placé devant des éléments de réalité langagière, devra avant tout faire preuve de capacité d'observation, les informations quantitatives étant fournies par les outils d'analyse informatiques. Cette approche est destinée à remplacer l'intuition, que l'apprenant ne possède d'ailleurs forcément que de manière limitée. Le premier cas présenté ici est celui d'un verbe relativement répandu: *entourer*, cherché à l'infinitif dans *CF* (annexe 1):

- Première constatation: *entourer* possède une très forte affinité avec le pronom *s(e)*, l'apostrophe étant son voisin de gauche le plus significatif; à l'infinitif, ce verbe est donc particulièrement utilisé sous sa forme pronominale *s'entourer*.
- Deuxième constatation: ce verbe a pour premier voisin significatif de droite la préposition *de* (sous ses formes *de*, *d'* et *des*). On repère donc une structure *s'entourer de*.
- Troisième constatation: parmi les cooccurrences significatives du verbe, on repère une très forte présence de formes du verbe savoir (*su*, *sait*, *sut*, *savoir*, *saura* et *savait*). On peut poser une structure *savoir s'entourer de*.
- Quatrième constatation: parmi les cooccurrences significatives, on trouve le pronom *il*. Ceci nous permet de retenir une structure *Il a su s'entourer de*.

Ce début de phrase *Il a su s'entourer de* demande maintenant à être développé. Intuitivement, on peut imaginer que des compléments appartenant à la catégorie [hum] pourront intervenir, ainsi qu'éventuellement des représentants d'autres catégories. Le retour au corpus permet de compléter le tableau en regroupant les termes dans les catégories sémantiques traditionnelles:

Il a su s'entourer de / d'...			
[hum]		[coll]	[abstr]
gens	bons	équipe	compétences
personnes	capables		précautions
amis	compétents		mystère
collaborateurs	expérimentés		
spécialistes	meilleurs		
experts			
professionnels			

Fig. 2: Complémentation de *s'entourer* de selon CF

On constate une forte prédominance des noms appartenant à la catégorie [hum], qui sont de surcroît très souvent qualifiés par des adjectifs de type mélioratif. Si la présence d'un terme collectif n'est pas surprenante dans ce contexte, en revanche le corpus livre quelques autres termes relevant du domaine abstrait. On remarque enfin parmi les voisins significatifs de droite d'*entourer* la présence de deux adverbes: *complètement* et *progressivement*. Une présentation condensée de ces informations donnera lieu à la phrase type suivante: *Il a su s'entourer [progressivement] de collaborateurs compétents*⁹.

L'observation de deux ressources écrites susceptibles d'être consultées par les apprenants (*Petit Robert* 2008 [PR] et *Dictionnaire des verbes du français actuel* 2010 [DVFA]) montre que l'emploi pronominal n'occupe pas une place prépondérante dans la présentation du verbe *entourer* (annexe 2). Les indications communes à ces deux ouvrages peuvent se résumer en trois points:

- *s'entourer de* [hum] avec une paraphrase "réunir autour de soi".
- *s'entourer de* [abstr] > *de confort*
- *s'entourer de* [abstr] > *de précautions*¹⁰

⁹ Il est très facile d'illustrer cet usage prototypique du verbe par une phrase authentique: *Ce médecin allemand a conduit ses recherches avec rigueur et esprit scientifique, il a su s'entourer de collaborateurs compétents* (<http://facemiroirame.forumsfree.org/t1083-radiesthesie-le-reseau-global-peyre-hartmann-reseau-h-19>).

¹⁰ On s'étonnera quelque peu du choix de Florea & Fuchs (2010: 102) de présenter les compléments [hum] sous le titre *se V de SN* et les compléments [abstr] sous *se V de N*, alors qu'il s'agit manifestement de la même structure syntaxique.

La figure 3 offre un aperçu comparatif des informations livrées par les trois instruments sollicités:

Sujet		PR	DVFA	CF
COI	[hum]	amis	<i>conseillers</i>	gens, personnes, amis , collaborateurs, spécialistes, experts, professionnels
	adjectif	—	meilleurs	bons, capables, compétents, expérimentés, meilleurs
	[coll]	—	<i>famille</i>	équipe
	[abstr]	<i>luxure, confort</i>	<i>confort</i>	—
		précautions	précautions	précautions , mystère
		—	—	compétences
	[concr]	<i>objets d'art</i>	<i>belles choses</i>	—
Autres	verbe	savoir	—	savoir , devoir, falloir
	adverbe	—	—	progressivement, complètement

Fig. 3: Complémentation de *s'entourer de*: *Petit Robert*, *Dictionnaire des verbes du français actuel* et *Corpus français* (en gras les éléments communs aux ouvrages écrits et à *CF*; en italique les éléments des dictionnaires qui ne se retrouvent pas dans *CF*)

Les résultats obtenus sont assez éloquentes et ne plaident pas franchement en faveur des dictionnaires. En position sujet, ceux-ci ne proposent pas d'argument, ne travaillant qu'avec des infinitifs; l'outil informatique de son côté ne propose qu'un pronom *il*, à vrai dire peu marqué. C'est donc en position complément indirect que les divergences entre les instruments traditionnels d'une part, *CF* d'autre part, sont les plus importantes:

- Seul un complément [hum] figurant dans les dictionnaires (*amis*) se retrouvent dans *CF*, qui en livre six autres, et surtout PR ne rend absolument pas compte des multiples adjectifs à valeur méliorative qualifiant ces noms.
- Concernant les compléments [coll], DVFA en présente un, *famille*, absent de *CF*. On notera que ce terme *famille* fait partie des cooccurrences significatives du participe passé *entouré*, ce qui démontre une valeur aspectuelle évidente: on *est* entouré de sa famille dans certaines situations de vie, on ne s'entoure pas de sa famille.
- Les termes de la catégorie [abstr] peuvent être classés en trois sous-catégories: si *luxure* et *confort* constituent un plus indéniable des deux

ouvrages, ces deux mots ne jouent qu'un rôle extrêmement marginal sur le plan quantitatif dans la réalité langagière; *précautions*, lui, est commun aux trois ressources; enfin, le terme *compétences* n'apparaît que dans *CF*, alors qu'il est en chiffres absolus le terme abstrait comptant le plus d'occurrences¹¹.

- Les compléments [concr] présentés dans les deux dictionnaires n'apparaissent pas dans la banque de données électronique.

C'est toutefois dans les cooccurrences significatives ne faisant pas partie des compléments du verbe au sens restreint que l'apport de *CF* se manifeste le plus clairement. En effet, outre le verbe *savoir* mentionné plus haut (également présent dans PR), on trouve les semi-auxiliaires *devoir* et *falloir*, qui donnent une idée plus précise des contextes d'utilisation de *s'entourer de*, qui comportent une notion de nécessité. La présence des deux adverbes *progressivement* et *complètement*, enfin, implique soit une valeur itérative – qui peut être liée à une notion d'effort ou du moins de travail régulier – ou au contraire une valeur résultative, comme point final d'un processus.

3.2 Désambiguïsation: regarder et ses (quasi-)synonymes¹²

À un niveau C1, l'une des principales tâches de l'apprenant est l'enrichissement de son vocabulaire dans la langue cible. Pour atteindre à une plus grande variété lexicale, il se voit ainsi régulièrement proposer des exercices portant sur les (quasi-)synonymes, soit des mots qui, selon la définition communément admise, ne sont interchangeables sans modifier notablement l'énoncé que dans certains contextes. La méthode traditionnelle, utilisée aussi bien dans l'enseignement de la langue maternelle que dans celui du FLE, vise à présenter ces parasyonymes comme faisant partie d'un même champ générique, pour ensuite insister sur les différences qui les opposent (Picoche, 1993: 41 ss.). Le recours à la paraphrase est également un procédé fréquemment utilisé pour sensibiliser l'apprenant aux différences de sens opposant les mots présentés (Blanchard et al., 1995: 159ss.). Si ces méthodes s'avèrent "efficaces", elles peuvent toutefois être considérées comme "insuffisantes" (Picoche, 1993: 51ss.) et peut-être également comme frustrantes pour l'apprenant, puisque le savoir est à chaque fois délivré par

¹¹ Une recherche sur Internet au moyen de Google le confirme: pour "s'entourer de luxe", on n'obtient que 349 résultats, pour "s'entourer de confort" 9'830. À titre de comparaison, on en trouve 22'400 pour "s'entourer de précautions" et 6'380 pour "s'entourer de mystère". Enfin, "s'entourer de compétences", grand absent des dictionnaires, apparaît 31'400 fois (état: 30 septembre 2011). À propos des recherches avec Google, voir les articles de Bickel et Deronne dans ce numéro.

¹² Je tiens à remercier ici mon ancienne étudiante, Mme Svetlana Baburina, auteure d'un travail de recherche empirique sur la synonymie.

une autorité externe, enseignant ou ouvrage de référence, et n'est pas le fruit d'une appropriation personnelle.

Le recours au corpus pourrait ici représenter une alternative didactique, puisqu'il permet de délimiter rapidement les contextes d'utilisation spécifiques de chacune des unités lexicales prises en compte; la désambiguïsation qui en résulte et les exemples authentiques illustrant les diverses utilisations livrés par la banque de données constituent un gain incontestable qui peut être exploité aussi bien par les enseignants que par les apprenants eux-mêmes. Nous allons exemplifier cette procédure de découverte en tentant de définir les contextes typiques d'utilisation des quasi-synonymes *contempler*, *dévisager*, *épier*, *scruter* et *toiser*, ainsi que du verbe hyperonyme *regarder*.

Le corpus livre une première indication purement quantitative qui n'est pas dénuée d'intérêt: le nombre d'occurrences de chaque verbe, dont nous avons recherché ici l'infinitif. Alors que l'hyperonyme *regarder* s'avère très fréquent (23'211 occurrences), *contempler* n'apparaît plus que 1'689 fois et *scruter* 720; quant aux verbes *épier* (195), *dévisager* (62) et *toiser* (58), ils peuvent être considérés comme d'un emploi rare¹³.

Les préférences lexicales de ces verbes nous en apprendront plus sur leur sens. Les listes de partenaires présentées ci-dessous ne se veulent pas exhaustives, mais sont représentatives des compléments et expansions enregistrés pour chacun des verbes.

3.2.1 Regarder (23'211)

L'hyperonyme *regarder* apparaît dans bon nombre de contextes fort différents. En fonction sujet, les deux catégories [hum] et [coll] sont les plus répandues, sans qu'un champ sémantique plus étroit ne se laisse réellement circonscrire: la plupart des termes possèdent en effet une valeur générale (*téléspectateurs*, *passants*, *monde*, *peuple*). Parmi les compléments directs, on constate une forte représentation de termes relatifs au domaine des médias, et particulièrement de la télévision (*télé*, *vidéo*, *film*), avec parfois une composante sportive supplémentaire (*match*, *jeu*). Deux domaines s'y ajoutent principalement: celui du spectacle naturel (*paysage*, *éclipse*) et, dans un sens plus figuré, celui de *l'avenir* et des *nouvelles perspectives*. Il n'est guère possible de rassembler les épithètes

¹³ Des recherches à la troisième personne du singulier (*regarde*, *contemple*, etc.) ou du pluriel (*regardent*, *contemplant*, etc.), ainsi qu'au participe passé (*regardé*, *contemplé*, etc.) confirment ce classement établi en fonction de la fréquence de chaque verbe. Elles montrent par ailleurs que les formes infinitives sont largement plus fréquentes que les autres (pour le verbe *regarder*, l'infinitif compte à lui seul autant d'occurrences que *regarde*, *regardent* et *regardé*).

collées à ces divers compléments sous une appellation générale, de dégager une fonction commune. Les adverbiaux présentant une affinité avec le verbe *regarder* expriment soit – dans un nombre restreint de cas – la qualité du regard (*avec confiance et foi, avec intérêt*), soit la proximité ou la perspective de celui-ci (*en arrière, en face, vers l'avant*). Les compléments supplémentaires jouent un rôle similaire à cette dernière catégorie: on regarde *sur un écran, sur un moniteur, etc.*

Sujet	COD	Adverbial
[hum] des centaines de curieux, les téléspectateurs, les amateurs de hockey, l'internaute, des passants [coll] le pouvoir, le basket français, le club de la Principauté, beaucoup de monde, le bon peuple	[abstr] le match, notre jeu, la deuxième chaîne, la télé, le concert, le film, des vidéos, les journaux une éclipse, le paysage, une ville, l'incendie les nouvelles perspectives, l'avenir les faits [concr] des sculptures, la montre	avec confiance et foi, avec intérêt, objectivement dans les yeux, de plus près, de trop près, en arrière, en face, vers l'avant, vers l'avenir sur plusieurs écrans dans le stade, sur son moniteur LCD, sur un grand écran

Fig. 4: Complémentation de *regarder* selon CF

En résumé, il s'avère difficile de déterminer un contexte réellement typique du verbe *regarder*. Cette constatation n'est à vrai dire pas étonnante dans la mesure où ce verbe faisant figure d'archilexème possède une valeur sémantique non-spécifique (voir Coseriu, 1970: 112), contrairement aux quasi-synonymes présentés ci-dessous.

3.2.2 Contempler (1'689)

Le sujet typique du verbe *contempler* est *un spectateur, un touriste, un amoureux ou un amateur*, donc un individu – ou un groupe d'individus – qui est venu exprès pour contempler, parfois longuement et attentivement, et qui y prend plaisir. Les compléments directs, généralement statiques, présentent en principe une valeur esthétique (spectacle de la nature ou artefacts à vocation artistique principalement); en outre, ces compléments sont fréquemment qualifiés par des épithètes à valeur méliorative, dont la

fonction est soit de renforcer l'idée de beauté, soit d'exprimer le caractère unique de l'objet. Enfin, les adverbiaux servent à mettre en évidence la durée et l'intensité de la contemplation, et expriment parfois l'admiration ressentie par le sujet.

Sujet	Verbe	COD	
[hum] le(s) spectateur(s), les touristes, un grand nombre de visiteurs les amoureux des astres, les amateurs, les nostalgiques [coll] le public	s'est contenté de ~, prennent plaisir à ~, s'attendrit de ~, passer du temps à ~, s'arrêter pour mieux ~, baver d'envie de ~, très touché de ~, s'attarder à ~	[abstr] le panorama, la nature, des paysages déserts, un coucher de soleil, le ciel étoilé, le paysage virtuel, le bain des petits oiseaux la beauté, le silence [concr] les photos, les œuvres d'arts, le tableau, les sculptures, la verroterie, l'image, le travail, des pages, le feu d'artifice, [hum] les femmes	belle, chouette, fantastiques, impressionnant, jolie, magnifique, majestueuse, sommptueuse, splendide, superbes mondialement célèbres et d'une valeur inestimable, originales, unique en son genre géants, obsolètes
Adverbial			
à loisir, tout mon soûl, pendant une vingtaine d'heures, longuement, le temps qu'il voudra		avec émerveillement, béatement, à l'aise	

Fig. 5: Complémentation de *contempler* selon CF

L'emploi prototypique du verbe *contempler* peut se résumer par les deux phrases suivantes: *Les amateurs prennent plaisir à contempler des œuvres d'art* ou *Les touristes s'arrêtent pour mieux contempler ce paysage magnifique*.

3.2.3 Scruter (720)

Le verbe *scruter* implique généralement un examen précis, la plupart du temps à des fins scientifiques, opéré par une personne qualifiée (*professionnel, photographe, astronome, etc.*), mais parfois également par

un *outil*, et s'étirant sur une certaine durée. Les compléments directs, presque toujours de nature abstraite, couvrent un large éventail de champs lexicaux, parmi lesquels on remarque notamment l'immense (*espace* ou *mer*) et le minuscule (*cristaux* ou *cellules*), ainsi que les termes relatifs au mouvement (*geste*, *déplacement*). Ces compléments sont parfois accompagnés de l'épithète *moindre*, qui renforce l'impression de concentration et d'effort soutenu, de même que les adverbiaux relativement fréquents (*attentivement*, *avec précision*, etc.). Le verbe *scruter* présente en outre une liste relativement longue de compléments supplémentaires: de moyen tout d'abord, avec l'expression des instruments de haute technologie utilisés (*télescope*, *laser*), ensuite de but (*trouver des solutions*, *rechercher Dieu*).

Sujet	Verbe	COD
<p>[hum] ces professionnels de prospection de météorites, les photographes et astronomes, les observateurs, le vérificateur, le sociologue</p> <p>[concr] cette panoplie d'outils, ses instruments, des outils informatiques particuliers, l'outil</p>	<p>passent leur temps à ~, avait passé des mois à ~, s'est ingénié à ~,</p>	<p>[abstr] les étoiles dans la Voie Lactée, les galaxies, les couches de l'atmosphère, les fonds marins</p> <p>la structure des cristaux, un grand nombre de cellules, le fonctionnement du cerveau, les «susceptibilités génétiques»</p> <p>le moindre frémissement, la moindre tache, chaque geste, les moindres déplacements, les moindres risées</p> <p>les colonnes de chiffres, les chiffres, l'immense base de Wikipedia</p> <p>tous les charivaris intérieurs des personnages, les ruines de l'âme, l'intérieur de soi</p> <p>le secret de Dieu</p> <p>[hum] l'être humain</p>
Adverbial		
<p>sous tous les angles, de près, avec la plus grande attention, pendant de longues minutes, de fond en comble,</p>	<p>avec précision</p> <p>soigneusement, constamment, plus attentivement</p>	
Complément de moyen		
<p>à travers un télescope, avec leur caméra, à la loupe,</p>	<p>au spectromètre de reconnaissance</p> <p>à l'aide de la technique,</p>	

avec la lunette, au laser,	à l'aide d'un logiciel, en utilisant les possibilités d'internet
Complément de but	
à la recherche d'aubaines, à la recherche de Dieu, à la recherche de la moindre trace, afin de déterminer, afin de tenter d'identifier le navire,	pour apprécier les différences, pour apprendre, pour éventuellement trouver des solutions, pour guetter l'apparition de la nouvelle lune

Fig. 6: Complémentation de *scruter* selon CF

Une présentation condensée des informations relatives à *scruter* est possible à travers les deux phrases types suivantes: *Des astronomes scrutent avec la plus grande attention la Voie lactée à travers un télescope* ou *Les chercheurs scrutent l'espace à la recherche d'autres formes de vie.*

3.2.4 Épier (195)

Le verbe *épier* exprime typiquement une action, présentée comme malhonnête, commise à l'insu de celui qui en est l'objet. Le sujet en est souvent une personne mal intentionnée (*malfaiteur, pédophile*) qui *se cache* ou *se terre* pour ses agissements. Les compléments directs sont de deux natures distinctes: [abstr] tout d'abord, avec *les (moindres) faits et gestes, les allées et venues, les déplacements*, [hum] ensuite, avec des personnes liées au sujet comme *ses voisins* ou *ses collaborateurs*, mais aussi des personnes inconnues comme *les jeunes filles* ou *des femmes seules*. Les compléments abstraits sont souvent accompagnés de *moindre* ou *tous*, qui renforcent l'idée que rien n'échappe à l'observateur. Les adverbiaux, de leur côté, insistent sur l'idée de secret et de dissimulation (*dans l'ombre, sans être repérée*, etc.), mais parfois aussi sur les efforts entrepris par le sujet et sur l'inconfort résultant de son activité (*grimpée dans un poirier, perché sur un arbre*). On notera enfin quelques compléments supplémentaires: de moyen (*avec des lunettes d'approche*) ou de but (*pour connaître, pour savoir*).

Sujet	Verbe	COD
[concr] l'objectif	se cache dans un buisson pour ~, avait pour tâche d'~, se terre dans un puits pour ~	[abstr] ses faits et gestes, les déplacements, tous mes gestes, les moindres faits et gestes, les allées et venues des délinquants, les déplacements des utilisateurs de téléphones mobiles GSM
[abstr] un logiciel malicieux, les programmes espions		les conversations téléphoniques
[hum] les malfaiteurs, le plaignant, le pédophile récidiviste		le sommeil des animaux

[coll] l'UMP et le PS		[hum] ses voisins, leurs adversaires, des passants, les jeunes filles dans leur dortoir, ses collaborateurs, des femmes seules et de tout âge, les uns les autres, le conjoint, les proches et les connaissances
Adverbial		
à votre insu, dans l'ombre, de plus près, grimpée dans un poirier,		perché sur un arbre, perché sur une branche d'olivier, sans être repérée, sans être vues
Complément de moyen		
avec des lunettes d'approche		
Complément de but		
pour connaître la vérité, pour des fins pseudo-scientifiques,		pour détecter les intrusions, pour savoir

Fig. 7: Complémentation de *épier* selon CF

En résumé, le verbe *épier* présente des usages prototypiques comme: *Le pédophile se cache dans un buisson pour épier les jeunes filles dans leur dortoir*, ou *Perché sur un arbre, ce citoyen épie les moindres faits et gestes de ses voisins avec des lunettes d'approche*.

3.2.5 Dévisager (62)

Le verbe *dévisager* exprime habituellement l'examen minutieux (*de très près, avec attention, sous toutes les coutures*) et parfois dissimulé (*sans lui en donner l'impression*) d'une personne par une autre. Un certain nombre de termes se retrouvent aussi bien en fonction sujet qu'en complément direct (*l'homme, les passants*), preuve que celui qui dévisage est susceptible d'être dévisagé à son tour. De plus, la présence de nombreuses occurrences à valeur passive (*se fait dévisager, en se faisant dévisager*) démontre que l'action peut se faire sans l'accord de l'objet, *discrètement*. Outre les compléments de la catégorie "être humain", on trouve un terme concret, *une voiture de sport*, objet de désir largement anthropomorphisé dévisagé par ses fans, ainsi que certains termes abstraits, dont des *sites internet* à vocation lobbyiste ou politique, qui se sont fait dévisager *par des cyberattaques arabes* ou *par le Mossad*. Il sera intéressant d'observer dans les années à venir si ces deux types de compléments gagneront en fréquence ou s'il ne s'agit que d'écarts accidentels par rapport à la norme.

Sujet	Verbe	COD
[hum] les fans, l'homme, un enfant, les passants, les spectateurs [coll] la foule de curieux	s'est fait ~, se fasse ~, en se faisant ~ par, passer son temps à ~, sans lui donner l'impression de le ~	[concr] la voiture de sport [abstr] cette ville alanguie, un site pro-islamique, les sites de lobbyistes [hum] les membres du groupe, les passants, les patients, les personnes, la jeune femme, l'homme
Adverbial		
avec plus d'attention, comme si elle était un monstre, de très près, mieux, radicalement,		sous toutes les coutures, très bien discrètement
Complément d'agent		
par le Mossad, par les jeunes attroupés dans la rue,		par des cyberattaques arabes

Fig. 8: Complémentation de *dévisager* selon CF

Une présentation condensée des informations sur le verbe *dévisager* s'exprimera dans des phrases comme: *Les fans dévisagent sous toutes les coutures les membres du groupe*, ou *Une femme se fait discrètement dévisager par un homme*.

3.2.6 Toiser (58)

Le verbe *toiser* apparaît en principe dans des contextes où il y a une opposition, une confrontation. Les acteurs typiques sont des êtres humains, parfois des termes collectifs, qui se toisent mutuellement, avec agressivité (*en chiens de faïence, méchamment*) ou avec dédain (*avec dégoût*). Le verbe est particulièrement utilisé dans le domaine sportif, où la rivalité est exacerbée, où il s'agit d'impressionner, d'intimider l'adversaire (en athlétisme, *le Jamaïcain toise les deux Américains*; en football, *les Marseillais toisent les Lionceaux*); cette tentative d'exercer par le regard un ascendant sur un concurrent se retrouve également dans les domaines politique et juridique.

Les compléments directs peuvent appartenir à la catégorie des termes abstraits, auquel cas la notion de regard mutuel disparaît plus ou moins complètement; de même, l'attitude de défi manifestée par les actants humains est absente des exemples suivants: *un pays n'aime plus guère*

toiser son passé communiste, une usine toise son avenir avec sérénité. Ce type d'emploi entraînant une désémantisation partielle du verbe *toiser* est pour le moins étonnant et demandera sans aucun doute à être examiné de près dans une perspective diachronique.

Sujet	COD	Adverbial
[concr] une dizaine de statues de soldats à taille humaine	[abstr] les ingratitude du monde, son passé communiste, son avenir, la question	avec cet humour corrosif, avec dégoût, du regard, en chiens de faïence, méchamment, mieux
[abstr] la justice	[hum] l'accusé, les rivaux, son rival, les deux Américains, l'Alsacien, son géniteur, son vis-à-vis, les Lionceaux	sereinement
[hum] le Jamaïcain, les candidats, les êtres humains, les magistrats, les Marseillais, Marathon Man, Mikko Hirvonen, fer de lance du team Ford	[coll] le public, son groupe F, le Real Madrid, l'assemblée	
[coll] Manchester, l'Angleterre		

Fig. 9: Complémentation de *toiser* selon CF

Les usages prototypiques de *toiser* peuvent se résumer dans les contextes suivants: *Les soldats toisent leurs ennemis avec dégoût*, ou *Le sportif toise son rival*.

4. Conclusion

Alors que le *data-driven learning*, sous forme de "totale prise en charge par les apprenants et du choix et de l'utilisation des documents authentiques" (Duda & Tyne, 2010: 86), est encore parfois considéré comme trop difficile, démotivant et inefficace par certains enseignants et apprenants, les deux procédures décrites ci-dessus nous semblent démontrer que cette pratique est désormais à la portée d'un apprenant ne disposant que de compétences limitées à la fois sur les plans de la maîtrise technologique et de l'analyse linguistique. Dans les domaines de la lexicologie et du lexique-grammaire, cette véritable interaction directe des apprenants avec le corpus (Boulton, 2010: 2), tenant compte non seulement des arguments d'un verbe au sens strict, mais également d'autres éléments présentant des affinités sémantiques avec le terme de recherche, peut en outre produire une qualité de résultats tout à fait remarquable en comparaison

d'outils traditionnels comme les dictionnaires ou les manuels. Avec un minimum de formation et d'information, il est ainsi possible de redéfinir la collaboration entre enseignant et apprenant, de renforcer l'individualisation de l'enseignement, l'autonomie de l'apprenant et, probablement, sa motivation.

Bibliographie

- Blanchard, S. et al. (1995): Vocabulaire. Paris (Nathan).
- Blanche-Benveniste, C. (2002): La complémentation verbale: petite introduction aux valences verbales. In: M.-J. Béguelin, J.-F. de Pietro & A. Näf (éds): Approches interlinguistiques de la complémentation verbale: quels savoirs pour l'enseignant? Quels savoirs pour l'élève? *TRANEL* 37, 47-73.
- Boulton A. (2008): Esprit de corpus: Promouvoir l'exploitation de corpus en apprentissage des langues. In: *Texte et corpus*, n°3, Actes des Journées de la linguistique de Corpus 2007, 37-46.
(disponible sur http://web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES_JLC07_boulton.pdf)
- (2010): Data-driven learning: on paper, in practice. In: T. Harris & M. Moreno Jaén (eds.): *Corpus Linguistics in Language Teaching*. Bern (Peter Lang), 17-52.
- Charaudeau, P. (2009): Dis-moi quel est ton corpus, je te dirai quelle est ta problématique. In: *Corpus* 8, 37-66.
- Church, K. & Mercer, R. (1993): Introduction to the Special Issue on Computational Linguistics Using Large Corpora. In: *Computational Linguistics* 1, 1-23.
- Coseriu, Eugenio (1970): Einführung in die strukturelle Betrachtung des Wortschatzes (Tübinger Beiträge zur Linguistik 14). Tübingen (Gunter Narr).
- Duda, R. & Tyne, H. (2010): Authenticity and Autonomy in Language Learning. In: *Bulletin suisse de Linguistique appliquée* 92, 86-106.
- Firth, J.R. (1957). *Papers in Linguistics 1934-1951*. Oxford (Oxford University Press).
- Florea, L.-S. & Fuchs, C. (2010): Dictionnaire des verbes du français actuel - Constructions, emplois, synonymes. Paris (Ophrys).
- Gross, M. (1975): *Méthodes en syntaxe*. Paris (Hermann).
- Harris, Z.S. (1976): *Notes du cours de syntaxe*. Paris (Le Seuil).
- Johns, T. (1991): Should you be persuaded: Two examples of data-driven learning. In: T. Johns & P. King (dir.): *Classroom concordancing*. *English language research journal*, Vol. 4, 1-16.
- Leech, G. (1991): The state of the art in corpus linguistics. In: K. Aijmer & B. Altenberg (eds.): *English corpus linguistics*. London (Longman), 8-29.
- Péry-Woodley, M.-P. (1995): Quels corpus pour quels traitements automatiques? In: *Traitement automatique des langues* 36 (1-2), 213-232.
- Picoche, J. (1993): *Didactique du vocabulaire français*. Paris (Nathan).
- Quasthoff, U., Richter M. & Biemann, C. (2006): Corpus Portal for Search in Monolingual Corpora. In: *Proceedings of the fifth international conference on Language Resources and Evaluation, LREC 2006, Genoa*, 1799-1802.
- Sánchez Cárdenas, B. (2010): Les restrictions sémantiques des arguments verbaux: une question de fréquence d'usage. In: *Synergies France* 6, 41-50.

- Rey-Debove, J. & Rey, A. (2008): *Le Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française*. Paris (Dictionnaires Le Robert).
- Saussure, F. de (1986): *Cours de linguistique française*. Paris (Payot). (1ère édition 1913).
- Sinclair, J. (1986): *First Throw Away Your Evidence*. In: G. Leitner (éd.): *The English Reference Grammar. Language and Linguistics, Writers and Readers (Linguistische Arbeiten 172)*. Tübingen (Niemeyer), 56-64.
- (1991): *Corpus, Concordance, Collocation*. Oxford (Oxford University Press).
- (2005): *Corpus and Text: Basic Principles*. In: M. Wynne (éd.): *Developing Linguistic Corpora: A Guide to Good Practice*, 1-16.
- Tesnière, Lucien (1959): *Éléments de syntaxe structurale*. Paris (Klincksieck).
- Tognini-Bonelli, E. (2001): *Corpus Linguistics at Work (Studies in Corpus Linguistics 6)*. Amsterdam (J. Benjamins).
- Williams, G. (2006): *La linguistique et le corpus: une affaire prépositionnelle*. In: F. Rastier & M. Ballabriga (dir.): *Corpus en Lettres et Sciences sociales — Des documents numériques à l'interprétation, actes du XXVII colloque d'Albi, Langages et signification*, 151-158. (<http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Williams.pdf>)
- (2008): *Introduction*. In: *Revue électronique Texte et corpus, n°3, Actes des Journées de la linguistique de Corpus 2007*, 7-14. (http://web.univubs.fr/corpus/jlc5/ACTES/ACTES_JLC07_introduction_williams.pdf).
- Willis, Dave (2003): *Rules, patterns and words*. Cambridge (Cambridge University Press).

Annexe 1: Résultats obtenus pour une recherche *entourer* dans *CF*.

W O R T S C H A T Z

UNIVERSITÄT LEIPZIG

Mot-clé: entourer

français

Rechercher



distinction entre minuscules et majuscules au début d'un mot

mot-clé: entourer

nombre d'occurrences: 1681

classe de fréquence: 14 (i.e., has got about 2¹⁴ the number of occurrences than the selected word.)

exemple(s):

Plus de 1800 personnes sont venues à la cathédrale Saint-Louis de Versailles pour entourer Pierre, Stéphane et Jacques-Bertrand, en ce jour si important pour eux. (source: *catholique-versailles.ccf.fr*)

Prendre les devants d'une enceinte : C'est l'entourer. (source: *voiletest.cher.tiscali.fr*)

L'auteur recommande aussi de former les anciennes patientes pour qu'elles puissent, à leur tour, aider à entourer les malheureuses jeunes mères qui sont souvent en outre victimes de rejet. (source: *www.cicred.med.fr*)
exemples supplémentaires

cooccurrences significatives de entourer:

s (253.91), su (909.13), ' (682.45), d (319.76), précautions (299.08), sait (295.67), compétents (245.12), bien (217.1), collaborateurs (197), qui (195.72), gens (193.24), de (173.23), garanties (168.45), et (165.2), il (164.49), sollicitude (132.58), pour (129.77), équipe (128.76), talent (109.4), conseils (104.42), mystère (96.43), aime (88.86), nmisticiens (86.46), (85.09), faut (84.84), continue (75.11), guillemets (73.64), m (73.55), personnes (72.53), soin (71.5), expérimentés (70.53), doit (66.21), spécialistes (65.66), jeunes (64.82), amait (63.78), consisie (62.47), sut (59.3), Pour (58.71), lui (56.59), savoir (56.14), qu (52.97), capables (50.64), protéger (49.05), leur (48.8), murailles (48.8), avis (48.59), historiens (46.43), les (46.43), vétérans (46.4), techniciens (45.98), meilleurs (45.47), experts (44.41), amis (44.31), Il (44.2), ou (44.16), toutes (43.76), S (43.15), devra (42.69), saura (42.24), dont (42.1), bons (41.86), savait (41.35), ses (40.91), compétentes (40.77), flou (40.35), maximum (40.01), toujours (40), afin (39.79), sensibilités (39.72), pourra (39.64), fossé (39.55), peut (39.5), une (39.47), compétences (38.94), professionnels (38.64), magouilleurs (37.67), expérience (37.16), fir (36.43), nécessaires (35.96), aussi (35.85)

voisins de gauche significatifs de entourer:

' (5105.71), pour (220.98), vous (90.82), fir (68.01), doit (60.79), à (54.87), venus (44.49), firent (37.82), va (36.51), vont (34.87), bien (33.22), doivent (30.48), nous (30.16), pouvaient (29.73), pouvez (26.37), texte (21.02), peut (19.17), devaient (18.35), su (17.59), venu (16.69), Pour (15.57), viennent (15.17), devrait (14.9), pouvait (13.9), semble (13.58), faire (13.46), mieux (12.3), fait (11.23), les (10.84), peuvent (10.34), pouvoir (6.43), donc (5.35)

voisins de droite significatifs de entourer:

de (1046.14), d (991.17), les (119.31), le (109.11), des (78.66), la (48.3), complètement (37.76), leurs (19.33), progressivement (19.24), sa (19.02), (17.94), (17.87), Marie (16.8), ce (14.4), toute (12.62), chaque (9.43), ces (9.42), leur (9.28), cette (8.88), (5.92), notre (4.59), son (4.34)

Annexe 2: Informations sur *entourer* dans le *Petit Robert* et le *Dictionnaire des verbes du français actuel*.

Le Petit Robert	Dictionnaire des verbes du français actuel
<p>1. Garnir (une chose) de qqch. qui en fait le tour, mettre autour de. <i>Entourer une ville de murailles</i> <i>Entourer d'une clôture, d'un mur</i> <i>Entourer d'un cercle, d'une bordure</i> <i>Entourer un président de gardes du corps</i> <i>Entourer qqn de soins, de prévenances, d'égards</i></p>	<p>V + SN <i>Les collines entourent la ville</i> > encercler, enserrer <i>Mille dangers nous entourent</i> > cerner, assaillir <i>Il faut ~ Jean pour l'aider à oublier.</i> > soutenir, épauler</p>
<p>2. (choses) être autour de (qqch., qqn) de manière à enfermer <i>Les murailles qui entourent la ville</i> <i>Une clôture entoure le jardin</i> <i>Les monts, les forêts qui entourent la vallée</i> <i>Le bonheur entourait cette maison tranquille</i> <i>Comme une eau bleue entoure exactement une île</i></p>	<p>V + SN de SN ~ <i>le jardin d'un mur</i> > clore, enclore ~ <i>le malade d'une couverture</i> > envelopper, enrrouler qqn dans ~ <i>qqn de ses bras</i> > enlacer qqn ~ <i>les noms d'un trait</i> > encercler, faire qqch autour de qqch</p>
<p>3. (personnes) se porter, se tenir tout autour de <i>Les soldats entourent la ville</i> <i>La police entoura les manifestants</i></p>	<p>V + SN de N ~ <i>ses parents d'affection</i>: > donner beaucoup d'affection à ses parents ~ <i>sa vie de mystère</i> > envelopper</p>
<p>4. (personnes ou choses) être habituellement ou momentanément autour de (qqn) <i>Les gens qui nous entourent</i> <i>Les dangers qui nous entourent</i></p>	<p>seV de SN <i>s'~ de sa famille; s'~ des meilleurs conseillers</i>: > réunir autour de soi qqn</p>
<p>S'ENTOURER DE (v. pron.) Mettre autour de soi. <i>S'entourer d'objets d'art, de luxe, de confort.</i> Réunir autour de soi. Savoir s'entourer d'amis. FIG. S'entourer de précautions.</p>	<p>seV de N <i>s'~ de confort; s'~ de belles choses</i>: > se doter <i>s'~ de précautions</i>: > être très vigilant</p>
<p>5. (personnes) s'occuper de (qqn), aider ou soutenir par sa présence, ses attentions <i>Ses amis l'entourent beaucoup depuis son deuil</i></p>	