

# Du dictionnaire de fréquence au lexique pour les apprenant·e·s de FLE : l'exemple des adjectifs / noms communs de personnes

Daniel Elmiger (Genève) et Alain Kamber (Neuchâtel)

---

## Abstract

In French morphology, there is a certain proximity of categories and inflectional models between adjectives and nouns (e. g. *responsable* (adj.) – *le/la responsable* (noun)). This is hardly reflected in FLE textbooks (French as a foreign language), which tend to present the grammatical categories and inflectional paradigms separately.

The topic of transcategorisation is investigated using a large electronic corpus (*Corpus français Leipzig*), which will be briefly described, as well as the frequency dictionary created on the basis of the same data. The empirical approach is exemplified by several words that are subject to a certain categorical variation; this addresses the question theoretically and suggests further didactical considerations.

---

## 1 Introduction

L'acquisition du lexique constitue, on le sait, un enjeu majeur de l'enseignement/apprentissage d'une langue étrangère. Et dans ce domaine, les questions se posant à l'enseignant·e sont nombreuses ; si depuis les travaux de Gougenheim dans les années 1950 et la publication de son *Français fondamental* (1958), le critère de la fréquence des mots s'est largement imposé, on s'accorde aujourd'hui, à l'heure des grands corpus électroniques, à affirmer qu'il ne suffit pas de présenter aux apprenants des listes de mots, mais que la mise en place de stratégies supplémentaires est indispensable (Boulton/Tyne 2014: 58). Pour utiliser le lexique correctement, les locuteurs et les scripteurs doivent en effet disposer de différents types d'informations complémentaires, d'ordre morphosyntaxique, sémantique et pragmatique. L'une d'entre elles est sans conteste la connaissance de la catégorie à laquelle – ou auxquelles – un mot appartient. Gougenheim recense ainsi, dans sa première liste de mots fréquents, notamment 380 substantifs, 200 verbes et 100 adjectifs. Mais que faire quand les frontières entre les catégories sont floues ou qu'elles sont perméables ? L'exemple suivant permet d'illustrer cette difficulté ; il s'agit de la traduction par une étudiante universitaire d'un extrait du roman *Die Gerettete Zunge* d'Elias Canetti, dans le cadre d'un cours de traduction allemand-français pour des apprenants de français de niveau B2 :

Der Großvater Canetti hatte sich selbst heraufgearbeitet, von einem betrogenen Waisenkind, das jung auf die Straße gesetzt wurde, hatte er es zwar zu Wohlstand gebracht, aber in den Augen

des anderen Großvaters blieb er **ein Komödiant und ein Lügner**. (Elias Canetti: *Die gerettete Zunge*)

Le grand-père Canetti avait réussi professionnellement, de l'orphelin trompé qu'on avait abandonné à un jeune âge, il avait réussi à faire fortune, mais aux yeux de l'autre grand-père, il était resté **(un) comédien et (un) menteur**.

Alors que le texte d'origine comporte deux substantifs (*ein Komödiant und ein Lügner*), on voit dans la traduction de l'étudiante une hésitation, explicitement exprimée par l'utilisation des parenthèses : nom – avec ou sans déterminant – ou adjectif ? Si, selon les dictionnaires usuels auxquels ont accès les apprenants, chacun des mots peut appartenir aux deux catégories grammaticales, il serait intéressant de mettre à leur disposition des outils propres à leur permettre de faire un choix motivé et pensé.

Nous nous intéresserons donc ici à un phénomène largement absent de la description grammaticale des manuels et ouvrages communément utilisés dans l'enseignement du FLE, celui de la « conversion » ou « distorsion catégorielle », et présenterons des moyens d'accéder à ces informations par une étude de corpus réalisée par l'enseignant·e-chercheur·e au profit de l'apprenant·e<sup>1</sup>.

Dans un premier temps, nous rappellerons quelques aspects théoriques de la conversion, en nous centrant plus spécifiquement sur les adjectifs / noms communs de personnes (point 2). Ensuite, nous présenterons le corpus et les outils utilisés pour accéder aux informations recherchées (point 3). Dans la partie empirique, nous présenterons quatre études de cas, liées à des procédés morphologiques différents : participes passés, participes présents, suffixation en *-able* et suffixation en *-eur* avec ses formes fléchies (point 4). Enfin, nous procéderons à une brève synthèse des résultats et ferons le bilan de la procédure dans la conclusion (point 5).

## 2 Théorie : flexion des noms et adjectifs

### 2.1 Proximité des catégories

La proximité formelle, sémantique et catégorielle entre adjectifs et noms a souvent été décrite dans la théorie grammaticale et a fait l'objet de nombreuses études. En l'absence d'un contexte syntaxique précis, il est souvent impossible, pour certains mots, de savoir s'ils s'utilisent comme adjectifs, comme noms ou si les deux catégories sont possibles. Et même en contexte, il n'est souvent pas simple de distinguer entre emploi adjectival et emploi substantif, comme dans les deux exemples suivants :

- 1) Cet accord est **créateur** d'emplois. (plutôt adjectif)
- 2) Il est **créateur** de bijoux. (plutôt nom)

Dans ce qui suit, nous allons brièvement esquisser les questions de la catégorisation, du sens de la conversion et des considérations sémantiques qui y sont liées.

---

<sup>1</sup> Si l'on se réfère à Boulton et Tyne (2014: 7), qui déterminent trois approches de l'apprentissage des langues au moyen de corpus, le travail présenté ici correspond plus spécifiquement aux études de type (b), soit une approche où les apprenants sont les *bénéficiaires* des données (utilisation « indirecte »), leur type (a) correspondant à la situation où les apprenants sont la *source* des données et leur type (c) à celle où les apprenants sont les *utilisateurs* des données (utilisation « directe »).

- Conversion, distorsion catégorielle ou transcatégorisation ?

La proximité catégorielle entre noms et adjectifs homonymes peut être décrite de différentes manières. Mel'čuk (1996: 23, 123) et Riegel et al. (2014: 908), par exemple, utilisent le terme de « conversion ». Pour Mel'čuk (1996: 123), celle-ci relève de la valeur syntaxique instable d'un lexème : « Nous appelons *conversion* une substitution élémentaire [au sens mathématique] qui est applicable à un syntactique morphologique et qui en fait un autre syntactique morphologique ». Lecolle (2012: 86) en revanche pose une condition morphologique supplémentaire :

[...] on ne parle de conversion que lorsque le lexème construit [acquiert] les propriétés d'identité catégorielle, pleine et nouvelle, que procure le statut de dérivé morphologique » (Kerleroux 1996: 189) : il doit donc posséder les caractéristiques syntaxiques de la catégorie d'arrivée [...]. Kerleroux (1996) parle ici, dans une approche syntaxique, de « distorsion catégorielle », ce qui signifie qu'il y a distorsion entre l'identité (ici la catégorie) du lexème et sa position syntaxique : *social*, pourtant adjectival, est employé comme un nom, mais avec des restrictions que n'ont pas les noms comparables que sont les noms de qualité.

(Lecolle 2012: 86)

Reprenant ainsi le concept de « distorsion catégorielle » développé par Kerleroux (1996), Lecolle (2011: 298) l'oppose à la « conversion » elle-même. Pour rendre compte de ces différents processus, elle propose un terme hyperonyme : « Le terme générique de « transcatégorisation » (ici adjectif/nom), englobant la conversion *stricto sensu* (comme fait de construction d'un nouveau lexème) et les phénomènes de distorsion catégorielle comparables, est employé pour rendre compte de cette continuité » (Lecolle 2011: 298).

Des phénomènes de type *société mère* ou *pack abonné*, où les noms *mère* et *abonné* s'utilisent pour spécifier un type de société ou de pack donné, ont également été décrits dans la littérature spécialisée, notamment par Noailly (1990), qui les décrit comme des « substantifs épithètes ». Comme dans le contexte de la distorsion catégorielle, le nom épithète ne prend pas nécessairement les marques formelles de la catégorie d'arrivée, en maintenant son autonomie de nom.

- Sens de la conversion et apport sémantique

S'il est souvent possible de retracer le sens de la conversion (adjectif dénominal ou nom déadjectival) en diachronie<sup>2</sup>, dans une perspective synchronique, en revanche, « le sens de la conversion est formellement inaccessible aux locuteurs » (Apothéloz 2002: 95). Dans cet article, nous ne chercherons pas à déterminer dans quel sens a pu s'opérer la conversion, en adoptant une position purement synchronique.

Le changement catégoriel (ou plus généralement : la transcatégorisation) est susceptible de donner lieu à une néologie catégorielle de type sémantique (cf. Lecolle 2011: 312–313, 2012) : « à partir du potentiel couvert par l'adjectif et l'adjectif nominalisé, la langue sélectionne et stabilise certains items dans certaines valeurs, en fonction de ce que « le monde » (social, culturel) retient. [...] Pourquoi *un documentaire* (film) et pas *\*un*

---

<sup>2</sup> Voir cependant les réserves émises par Lecolle (2011: 298), qui constate que « dans nombre de ces cas, il est difficile, voire illusoire, de déterminer l'antériorité chronologique d'une catégorie par rapport à l'autre. »

*littéraire* (production écrite) ? ». Dans cet exemple comme dans bien d'autres cas, le passage d'une catégorie à l'autre peut ainsi donner lieu à des différences sémantiques imprévisibles.

## 2.2 Différences des formes flexionnelles

Du point de vue morphologique, on trouve des paradigmes similaires pour les noms communs de personnes et pour les adjectifs. Ceci est d'autant plus vrai pour les mots utilisés dans les deux catégories lexicales. Le tableau 1 donne une vue d'ensemble simplifiée des principaux procédés morphologiques que l'on peut observer :

	ADJ-m	ADJ-f	NOM-m	NOM-f
hétéronymie			<i>le frère</i>	<i>la sœur</i>
suffixe			<i>le poète</i>	<i>la poétesse</i>
désinences différentes	<i>(plan) directeur</i>	<i>(ligne) directrice</i>	<i>le directeur</i>	<i>la directrice</i>
formes épiciènes (masc. = fém.)	<i>(avis) responsable</i>	<i>(attitude) responsable</i>	<i>le responsable</i>	<i>la responsable</i>

**Tableau 1 : exemples de formes flexionnelles adjectivales et nominales**

La question de savoir si ces paradigmes relèvent de la flexion ou de la dérivation est une question morphologique, mais aussi idéologique, particulièrement en ce qui concerne les noms communs de personne (cf. Elmiger 2013). Elle touche un très grand nombre d'adjectifs et de noms (communs de personne), notamment ceux qui se différencient par un *-e* au féminin (par ex. *indépendant-e*, n. et adj.). Traditionnellement, on range les adjectifs dans le domaine de la flexion (en principe, les formes sont accordées selon le nombre et le genre du noyau (ou « donneur » de genre) ; pour les noms communs, en revanche, on privilégie souvent une explication par dérivation, moins contraignante, mais aussi moins régulière (pour les différences entre flexion et dérivation, cf. Haspelmath 2010: 89).

Il n'est pas toujours aisé de distinguer entre emploi adjectival et emploi nominal – d'autant plus que ce ne sont pas les seules catégories pertinentes. Le tableau 2 montre, pour quelques mots, leur distribution dans les catégories lexicales auxquelles ils peuvent théoriquement appartenir :

	adjectif	participe	nom commun de chose	nom commun de personne	gentilé <sup>3</sup>
<i>abonné-e</i>	x	x		x	
<i>accusé-e</i>	x	x	x <sup>4</sup>	x	
<i>acheteur, acheteuse</i>	x <sup>5</sup>			x	
<i>acteur, actrice</i>	(x)			x	
<i>adulte</i>	x			x	
<i>africain-e</i>	x				

<sup>3</sup> Nom que portent les habitant·e·s d'un lieu.

<sup>4</sup> Avec le sens : 'avis, notification' (*accusé de réception*).

<sup>5</sup> Cf. par exemple le syntagme *fièvre acheteuse*.

<i>Africain·e</i>					x
<i>bleu·e</i>	x		x	x <sup>6</sup>	
<i>Bleu·e</i>				x	x
<i>allié·e</i>	x	x		x	

**Tableau 2 : appartenance théorique de quelques mots-formes à différentes catégories lexicales**

Le tableau 2 appelle quelques commentaires. Premièrement, il n'est pas certain que toutes les possibilités correspondent à des usages effectifs et réguliers. Dans le cas de *abonné* et *accusé*, l'emploi adjectival et nominal (en tant que nom commun de personne) ne fait intuitivement aucun doute ; quant à la valeur adjectivale de *acheteur/acheteuse*, on peut observer une collocation forte pour *fièvre acheteuse* (mais il n'est pas certain que ce mot ait une valeur adjectivale dans d'autres contextes). En revanche, l'emploi adjectival d'un mot comme *acteur/actrice* peut quant à lui être soumis à caution.

Ensuite, on peut constater, au niveau de la forme graphique, que la majuscule, en graphie standard, est utilisée pour différencier entre l'usage adjectival (*africain·e*) et l'usage nominal gentilé (*Africain·e*).

Enfin, on constate qu'un mot comme *acteur* peut désigner aussi bien un être humain qu'un terme collectif (entité administrative, économique ou juridique), comme par exemple dans *Pierre est un acteur / la firme est un acteur*.

### 3 Les outils et leur description

Le moyen le plus adapté pour passer des possibilités théoriques de variation catégorielle d'un mot à sa répartition effective sur plusieurs catégories grammaticales est sans conteste de procéder à une recherche quantitative au moyen d'un corpus et d'un concordancier. Ainsi que le rappelle Williams (2006: 152), dans la tradition anglo-saxonne de la linguistique de corpus, et particulièrement dans l'école contextualiste de Birmingham représentée par Sinclair, « la lexicographie, l'enseignement et les corpus sont intimement liés ». Cet apport de la linguistique de corpus au développement de la lexicographie se manifeste principalement dans deux types d'approche : d'une part l'étude des colligations et des collocations<sup>7</sup>, d'autre part des approches intégrant une dimension plus spécifiquement morphologique, souvent liées au traitement automatique du langage (TAL)<sup>8</sup>.

Le présent travail se fonde sur une vaste banque de données qui a donné lieu à deux instruments qui seront brièvement décrits dans les parties 3.1 (*Corpus français Leipzig*) et 3.2 (*Dictionnaire de fréquence du français*). Ces deux outils traitent des données recueillies par une équipe de recherche de l'Université de Leipzig et constituent à notre avis d'excellents instruments pour la recherche sur corpus à des fins diverses (usage quotidien, recherche scientifique, utilisation en contexte scolaire, etc.).

<sup>6</sup> Avec le sens : 'débutant'.

<sup>7</sup> Pour l'anglais, voir bien entendu les travaux de Sinclair (et particulièrement Sinclair 1991), pour le français, notamment Tutin et Grossmann (2003) et Blumenthal, Novakova et Siepmann (2014), pour l'allemand, Steyer (2013).

<sup>8</sup> Cf. ten Hacken et Panocová (2013), divers travaux de Dal et de Hathout (cf. bibliographie) ; pour l'allemand, se référer à Hass (2005) et Hahn et al. (2008).

### 3.1 *Corpus français Leipzig*

Il existe différents types de corpus, qui sont construits tantôt en fonction d'une thématique de recherche particulière (par exemple *Emolex*<sup>9</sup>) ou d'un genre textuel spécifique (par exemple *Scientext*<sup>10</sup>), tantôt pour permettre certaines formes limitées de recherches, notamment lexicographiques. *Corpus français Leipzig* (désormais *CFL*) fait incontestablement partie de ce deuxième type. Contrairement au *Corpus de référence du français contemporain* (*CRFC*), qui a vocation d'être équilibré et dont les 310 millions de mots actuels se répartissent également sur l'oral (avec une sous-catégorie *pseudo-oral* comptant notamment des sous-titres de films et des discussions sur des forums) et l'écrit et couvrent une période allant de 1945 à 2014 (voir Siepmann 2015: 143), *CFL* est un corpus général de la langue française écrite contemporaine qui – à l'instar de *Frantext*<sup>11</sup> par exemple qui est avant tout un corpus littéraire – n'est pas un corpus équilibré puisqu'il contient surtout des écrits journalistiques, *Wikipedia* et divers sites Web depuis les années 2000. Sa taille est assez importante, avec ses 1,46 milliard de mots<sup>12</sup> (en comparaison avec d'autres langues comme l'anglais ou l'allemand, où les grandes collections comptent des milliards de mots, les corpus du français gardent en général une taille relativement modeste) et permet de tirer des enseignements intéressants en matière de fréquences de mots.

Le *CFL* fait partie d'une série de corpus différenciées selon la langue (ou une variété régionale) : en tout, 238 collections monolingues (*Corpus-Based Monolingual*), de l'abkhase au zoulou. La présentation des informations peut varier d'un corpus à l'autre, mais en général, les informations suivantes sont données<sup>13</sup> :

- le mot-clé, c'est-à-dire la chaîne de signes recherchée ;
- le nombre d'occurrences trouvées dans le corpus ;
- des indications sur la fréquence du mot, notamment un indice de fréquence (relative à celle du mot le plus fréquent : *de*) ;
- exemples : cinq phrases contenant le mot-clé apparaissent automatiquement ; d'autres exemples sont disponibles ;
- cooccurrences : liste des mots ayant un indice de cooccurrence élevé par rapport au mot-clé recherché ;
- voisins de gauche / de droite significatifs : liste des mots qui se trouvent le plus souvent dans le voisinage immédiat du mot-clé ;

<sup>9</sup> <http://emolex.eu/> (corpus pour l'étude du lexique des émotions dans cinq langues européennes : allemand, anglais, espagnol, français et russe).

<sup>10</sup> <http://scientext.msh-alpes.fr/> (corpus pour l'étude du positionnement et du raisonnement de l'auteur dans les écrits scientifiques : français et anglais).

<sup>11</sup> 277 413 739 mots, du X<sup>e</sup> au XXI<sup>e</sup> siècle ([www.frantext.fr](http://www.frantext.fr)).

<sup>12</sup> Cf. <http://corpora.informatik.uni-leipzig.de> (sélectionner « French »). Depuis juin 2015, la recherche se fait via une nouvelle interface ; le nombre de mots accessibles pour le français a doublé par rapport à la version précédente.

<sup>13</sup> Pour une description plus détaillée de *CFL*, cf. Kamber (2011, 2014).

- un graphe illustrant les liens de fréquence à l'aide d'un réseau de correspondances autour du mot-clé.

- **Utilisation des ressources hors ligne**

Pour des raisons techniques, mais aussi de droits d'auteur, l'utilisation de *CFL* en ligne est limitée de deux manières : d'une part, le niveau structurel le plus élevé accessible est celui de la phrase (graphique), ce qui limite l'intérêt de l'outil pour tous les phénomènes transphrastiques et textuels. D'autre part, la recherche en ligne est limitée à un seul mot-clé, ce qui empêche des recherches plus complexes. Cette deuxième limitation peut être contournée par une utilisation hors ligne des ressources : une partie des données peut être téléchargée pour l'utilisation dans divers logiciels de concordance (en tant que texte brut ou données MySQL) : différentes sélections de textes sont à disposition, allant de 10 000 à 10 millions de phrases (c'est-à-dire quelque 200 millions de mots, <http://corpora.informatik.uni-leipzig.de/download.html>, [20.05.2015]).

### 3.2 Le *Dictionnaire de fréquence du français*

Le *Dictionnaire de fréquence du français* (Eckart et al. 2013) fait partie de la série *Frequency Dictionaries*, qui a pour objectif de livrer des indications de fréquence comparables pour un grand nombre de langues différentes<sup>14</sup> grâce à une procédure de traitement unifiée des ressources électroniques de la *Leipzig Corpora Collection* (<http://corpora.informatik.uni-leipzig.de/>, [20.05.2015]). La version papier du dictionnaire contient une description des données et de l'approche méthodologique et fournit des informations statistiques sur l'utilisation des lettres, la structure des mots et les changements structurels du lexique en fonction des classes de fréquence. Elle présente les 1 000 formes lexicales les plus fréquentes classées par ordre de fréquence et les 10 000 formes les plus fréquentes par ordre alphabétique. Une vérification à la main a permis d'y identifier les formes erronées (par ex. *\*etre*) ou correspondant aux Rectifications orthographiques de 1990 (par ex. *boite*). Le CD-ROM en annexe contient en outre une liste d'un million de formes classées par ordre alphabétique ou par ordre de fréquence. Cette liste plus étendue n'a été contrôlée qu'au moyen de critères automatiques.

Les listes du *Dictionnaire de fréquence* présentent des formes lexicales (ou mots-formes au sens de Mel'čuk 1993: 99) et non des lexèmes ou unités lexicales<sup>15</sup>. Cela entraîne trois conséquences :

1. Tout d'abord, l'absence de mots composés tels que *pomme de terre*, un lexème formé de trois mots-formes,
2. Ensuite, le besoin de chercher toutes les formes fléchies d'une même unité lexicale dans la banque de données pour avoir une vue d'ensemble du fonctionnement de celle-ci (par ex. 38 formes pour un verbe régulier en *-er*<sup>16</sup>),

---

<sup>14</sup> Actuellement, cette série compte, outre le français, des volumes dédiés au lexique de l'allemand, de l'anglais, de l'islandais, du hongrois, de l'espéranto, de l'indonésien et de l'ukrainien.

<sup>15</sup> Sur les différents choix auxquels ont procédé les auteurs pour définir la notion de mot, voir Eckart et al. (2013: 6–8) ainsi que Quasthoff et al. (2014).

<sup>16</sup> Sur cette question, voir Kamber (2011).

3. Enfin, la nécessité de désambiguïser les mots-formes à la main, ceux-ci pouvant être attribués à différents lexèmes pour différentes raisons, notamment :

- les formes fléchies sont rattachées à deux unités lexicales différentes (*sommes* peut être une forme conjuguée du verbe *être* ou le pluriel du nom *somme*),
- les formes de base elles-mêmes sont ambiguës (*que* peut correspondre à la conjonction ou au pronom relatif),
- la forme de base est identique et polysémique (*voler* est un verbe ayant deux sens différents).

Les points (2) et (3) jouent un rôle primordial dans cette étude. En effet, pour pouvoir dresser un tableau exhaustif de la répartition d'un mot sur les différentes catégories grammaticales, il sera nécessaire de tenir compte de toutes les formes fléchies de celui-ci en fonction du genre et du nombre, et de définir des modalités pour procéder à la main au tri des occurrences enregistrées.

#### 4 Recherche dans *Corpus français Leipzig et résultats*

Dans le cadre de cette étude exploratoire, nous nous sommes basés sur la liste des 10 000 mots les plus fréquents répertoriés dans le *Dictionnaire de fréquence* (Eckart et al. 2013). Ainsi, nous pensons que les cas que l'on peut y observer correspondent au lexique pertinent pour un public d'apprenant·e·s (rappelons que 10 000 mots-formes équivalent à un nombre bien inférieur de lexèmes)<sup>17</sup>. Dans cette liste, nous avons sélectionné, parmi les mots commençant par A, B ou C, quatre cas de figure susceptibles de présenter des distorsions catégorielles, particulièrement entre les catégories adjectifs / noms communs de personne : formes dérivées du participe passé, formes dérivées du participe présent, formes à suffixe *-able* et formes à suffixe *-eur*. Dans chacun des cas, nous avons tenu compte des distinctions masculin/féminin et singulier/pluriel, ainsi que, pour les formes en *-eur*, des différents processus morphologiques à l'œuvre en matière de féminisation.

Les données complètes de *CFL* n'étant pas disponibles pour une recherche à l'aide d'un concordancier, il faut se servir des sous-corpus mis à disposition pour une utilisation *offline*. Pour cette recherche, nous avons choisi de travailler à l'aide du plus grand corpus mixte téléchargeable disponible (« fra\_mixed\_2009\_10M-text »), présentant, à une échelle réduite, les mêmes caractéristiques que *CFL*, soit un ensemble composé de trois parties (journaux francophones, pages web et *Wikipedia*) pour 10 millions de phrases (près de 200 millions de mots). À l'aide du concordancier *AntConc*<sup>18</sup>, nous avons procédé à une recherche pour chaque forme fléchie des termes retenus, puis à une sélection aléatoire de 200 occurrences quand leur nombre dépassait ce seuil, qui ont ensuite été analysées à la main. Ce nombre n'est pas défini par les exigences d'une analyse statistique inférentielle, mais plutôt par des contraintes liées à l'analyse empirique, qui ne saurait être automatisée<sup>19</sup>. Cette utilisation d'un échantillon

<sup>17</sup> Gougenheim (1958) et *LexTutor* ([www.lextutor.ca](http://www.lextutor.ca)) se basent sur un nombre comparable de mots.

<sup>18</sup> Voir [www.laurenceanthony.net/software/antconcl/](http://www.laurenceanthony.net/software/antconcl/) (version utilisée : AntConc 3.4.3.0).

<sup>19</sup> Les analyseurs automatiques morphosyntaxiques disponibles actuellement ne sont pas toujours suffisamment fiables pour l'analyse de formes qui ont un statut catégoriel instable. Ainsi, l'analyse d'un outil comme *Treetagger* ([www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/treetagger.html](http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/treetagger.html)) s'avère souvent fautive :



aléatoire simple a été pratiquée dans d'autres études (p. ex. déjà chez Coates 1983) et présente l'avantage de rendre maniable une grande masse de données lorsque chaque occurrence doit être interprétée individuellement, tout en gardant un caractère largement représentatif de l'ensemble ; l'inconvénient majeur de la méthode, qui implique que seuls les phénomènes ayant une certaine fréquence seront présents dans l'échantillon réduit, ne portera pas à conséquence dans le cas de notre étude, où le choix est limité à deux, parfois trois classes grammaticales différentes. De la sorte, il est possible d'obtenir des nombres relatifs (en tant que pourcentages), ce qui permet la comparaison entre les formes étudiées, en laissant volontairement de côté la question de la fréquence absolue des différents cas.

#### 4.1 Les formes participiales passées

Pour les formes dérivées d'un participe passé, cinq cas ont été analysés : *abonné*, *accusé*, *associé*, *assuré* et *condamné*.

terme de recherche	abonné	abonnée	abonnés	abonnées	total formes
nbre d'occurrences (nbre total)	200 (1 362)	80	200 (1 812)	79	559 (7 336)
nom de personne	150	13	193	29	<b>385 (68,9%)</b>
participe passé	50	67	7	50	<b>174 (31,1%)</b>
passif	36	34	3	22	95 [54,6%]
valeur adjectivale	14	27	4	23	68 [39,1%]
s'abonner	-	6	-	5	11 [6,3%]

terme de recherche	accusé	accusée	accusés	accusées	total formes
nbre d'occurrences (nbre total)	200 (12 712)	200 (1 378)	200 (5 009)	200 (582)	800 (19 681)
nom de personne	37	32	103	12	<b>184 (23%)</b>
nom de chose	11	-	3	-	<b>14 (1,8%)</b>
participe passé	152	168	94	188	<b>602 (75,2%)</b>
passif	52	90	43	84	269 [44,7%]
actif /sens 1	59	10	-	-	69 [11,5%]

par exemple, *coupable* est identifié comme nom dans l'exemple suivant, tandis que le mot a clairement une fonction adjectivale : *Le droit actuel ne permet plus non plus d'éviter que les fournisseurs [...] soient tenus pour responsables des atteintes aux droits d'auteur dont se rend **coupable** leur clientèle.* (exemple tiré du corpus *Feuille fédérale*, cf. Elmiger 2015). Voir également Kübler (2014) sur les importants problèmes d'étiquetage du corpus en ligne *Les Voisins de le Monde* (<http://redac.univ-tlse2.fr/voisinsdelemonde>, [20.05.2015]).

actif / sens 2	11	-	-	-	11	[1,8%]
valeur adj. / sens 1	28	66	50	97	241	[40,0%]
valeur adj. / sens 2	2	1	-	1	4	[0,7%]
valeur adj. / sens 3	-	-	-	3	3	[0,5%]
s'accuser	-	1	1	3	5	[0,8%]

terme de recherche	associé	associée	associés	associées	total formes	
	200 (7 348)	200 (3 732)	200 (5 894)	200 (2 305)	800	(19 279)
nom de personne	27	2	36	0	65	(8,1%)
participe passé	173	198	164	200	735	(91,9%)
passif	54	73	48	60	235	[32,0%]
actif	4	-	-	-	4	[0,5%]
valeur adjectivale	106	118	109	138	471	[64,1%]
s'associer	9	7	7	2	25	[3,4%]

terme de recherche	assuré	assurée	assurés	assurées	total formes	
	200 (16 845)	200 (4 477)	200 (2 755)	200 (845)	800	(24 922)
nom de personne	11	2	94	1 [erreur]	108	(13,5%)
participe passé	189	198	106	199	692	(86,5%)
passif	33	146	54	120	353	[51,0%]
actif	140	2 [2 erreurs]	2	3 [1 erreur]	147	[21,2%]
valeur adjectivale	13	43	42	75	173	[25,0%]
s'assurer	3	7	8	1 [1 erreur]	19	[2,8%]

terme de recherche	condamné	condamnée	condamnés	condamnées	total formes	
	200 (11 965)	200 (1 565)	200 (3 858)	200 (667)	800	(18 055)
nom de personne	15	0	47	3	65	(8,1%)
participe passé	185	200	153	197	735	(91,9%)
passif	101	137	103	104	445	[60,5%]

actif	57	7	2	3 [1 erreur]	69 [9,4%]
valeur adjectivale	27	55	48	90	220 [29,9%]
se condamner		1			1 [0,2%]

Tableau 3 : distribution de cinq formes participiales passées

On note immédiatement, parmi les cinq cas observés, le profil particulier d'*abonné*, qui est le seul à apparaître majoritairement comme nom (plus de deux tiers des occurrences), alors que les quatre autres sont très largement des participes passés (entre trois quarts et plus de 90% des occurrences). Ces emplois nominaux sont particulièrement fréquents dans les formes masculines (respectivement 150 et 193 sur 200 au singulier et au pluriel), alors que les formes féminines, moins représentées dans le corpus (80 occurrences pour le singulier, 79 pour le pluriel), sont avant tout des participes passés :

- 3) Rappelons que pour engager une procédure de portabilité du numéro, **l'abonné** doit être libre de tout engagement.
- 4) La lecture de ce texte est réservée à **nos abonnés** en ligne.

La prédominance des emplois nominaux – au masculin – s'explique bien entendu en bonne partie par la constitution du corpus, et notamment par la présence de textes journalistiques et d'extraits de sites Internet, dans lesquels les moyens de communication électroniques sont largement présents : dans l'exemple (3), il s'agit d'un abonné aux services d'une entreprise de télécommunication, dans l'exemple (4) d'abonnés à la version en ligne d'un quotidien. On notera également la présence, parmi ces emplois nominaux, de la locution *aux abonné-e-s absent-e-s* (7 cas au total, dont 3 pour le féminin) :

- 5) La défense italienne fut également aux abonnées absentes sur l'essai portugais inscrit par Penalva, joueur de Blagnac (promu en Pro D2).

En ce qui concerne les formes participiales passées (majoritaires au féminin), on constate une large prédominance des formes passives<sup>20</sup> à valeur résultative (6), devant la valeur adjectivale (7), les formes pronominales (8) ne jouant, pour cette forme comme pour les autres d'ailleurs, qu'un rôle marginal<sup>21</sup> :

- 6) Ainsi, 7,4 millions de personnes **sont abonnées** à ce type d'accès, en hausse de 8,5% par rapport aux trois premiers mois de l'année.
- 7) Mais l'équipe, **abonnée** aux play-offs, n'accède jamais aux finales.
- 8) Quinze mille personnes se sont inscrites à l'essai gratuit, 1000 **se sont abonnées** par la suite.

Si la forme *accusé* présente trois quarts de formes participiales passées, elle n'en est pas moins remarquable dans son emploi nominal (présent particulièrement dans le masculin pluriel). En effet, elle présente un cas de distorsion catégorielle plus complexe que les autres formes analysées, puisqu'à côté d'un nom de personne – 184 occurrences sur un total de 800

<sup>20</sup> N'ont été considérées comme des formes passives dans cette étude que des cas où le participe passé apparaît explicitement en relation avec un auxiliaire *être*.

<sup>21</sup> Peut-être faut-il voir dans cet état de fait, constaté pour tous les verbes potentiellement pronominaux analysés ici, une conséquence de stratégies d'évitement de la part des scripteurs, l'accord des participes passés des verbes pronominaux constituant on le sait une difficulté majeure de la grammaire française.

analysées – (9), on recense un nom de chose dans les composés *accusé(s) de réception* (13 occurrences) – voir (10) – et *accusé(s) de dépôt* (1 occurrence) :

9) **Les accusés**, à l'exception de Garry Gabriel, sont en liberté.

10) Cela fait plusieurs mois, depuis septembre, que nous lui avons fait des demandes de ces pièces justificatives, nous possédons **des accusés de réception**.

Une autre particularité de cette forme est de présenter deux sens différents dans son emploi à l'actif – l'acception principale 'rendre responsable' (11), mais aussi 'subir' (12) dans un nombre restreint d'occurrences – et trois dans son utilisation adjectivale – à côté de 'rendu responsable' (13), également 'subi' (14) et 'marqué, prononcé' (15) :

11) La Russie **a accusé**, samedi 6 janvier 2007, les Etats-Unis d'imposer illégalement des sanctions à certaines entreprises d'armement russes qui, selon Washington, coopèrent avec l'Iran et la Syrie.

12) Le compte des investissements directs **a accusé** un nouveau déficit de 727,1 milliards, indiquant qu'en novembre, les Japonais ont encore investi à l'étranger plus que les étrangers n'ont investi au Japon.

13) Le rapport vise particulièrement les forces armées éthiopiennes, **accusées** de "riposter aux attaques rebelles par des tirs indiscriminés d'artillerie et de roquettes lourdes, avec un effet dévastateur sur les civils".

14) Les pertes **accusées** en soldant ces positions effaceront la plus grande partie des bénéfices attendus en 2007.

15) Mais surtout, le bouillon de crevettes est très bon, jouant toujours sur des saveurs **accusées**.

Comme pour les autres cas analysés, on note la prédominance du passif dans les emplois verbaux<sup>22</sup>, ainsi que l'utilisation très restreinte de la forme pronominale *s'accuser*.

Les trois dernières formes analysées, *associé*, *assuré* et *condamné*, présentent des profils très semblables. En effet, leur emploi nominal est très rare (entre 8,1% et 13,5% seulement), et on constate une quasi-absence de celui-ci au féminin. À chaque fois, ce sont les formes du masculin pluriel qui livrent, dans des proportions diverses, l'essentiel des noms :

16) Passionnés des matières, **les associés** continuent de développer la recherche des produits naturels et innovants.

17) Cela serait incohérent et complexe pour **les assurés**.

18) Selon **l'un des condamnés**, les garagistes auraient dans certains cas été complices des vols.

On notera encore, pour *associé*, une tendance modeste (4 occurrences) à recourir à la composition, comme dans :

19) Courtisé par plusieurs maisons, dont certaines anglo-saxonnes, il rejoint finalement Rothschild & Cie en tant qu'**associé gérant**.

Pour ce qui est des participes passés, on retrouve, dans des proportions variables, les quatre catégories citées ci-dessus pour *accusé*, soit le passif, l'actif, la valeur adjectivale et le verbe pronominal.

<sup>22</sup> La forme *associé* faisant ici figure d'exception.

## 4.2 Les formes participiales présentes

Pour illustrer les disparités en matière de distorsion catégorielle en relation avec les formes dérivées d'un participe présent, les cas d'*attaquant* et de *correspondant* ont été retenus.

terme de recherche	attaquant	attaquante	attaquants	attaquantes	total formes	
nbre d'occ. (nbre total)	200 (10 617)	113	200 (1 893)	29	542	(12 652)
nom	190	94	200	22	506	(93,4%)
adjectif	-	19	-	7	26	(4,8%)
participe présent	10	-	-	-	10	(1,8%)

terme de recherche	correspondant	correspondante	correspondants	correspondantes	total formes	
nbre d'occ. (nbre total)	200 (8 415)	200 (1 559)	200 (2 219)	200 (704)	800	(19 681)
nom	63	32	107	3	205	(25,6%)
adjectif	37	168	93	197	495	(61,9%)
participe présent	100	-	-	-	100	(12,5%)

Tableau 4 : distribution de deux formes participiales présentes

Les deux formes participiales présentes sélectionnées présentent des profils très différents. Cela commence par leur nombre d'occurrences dans le corpus, les formes d'*attaquant* étant rares au féminin, et se poursuit par leur répartition sur les différentes catégories grammaticales.

Pour *attaquant*, la prédominance des emplois nominaux est écrasante, indépendamment du genre et du nombre (93,4% du total), avec notamment un emploi exclusivement nominal pour le masculin pluriel<sup>23</sup> :

20) Les **attaquants** bourguignons en avaient visiblement marre de passer pour des pieds carrés.

L'emploi adjectival, réservé au féminin largement moins représenté dans le corpus, ne représente que 4,8% du total :

21) Au signal du maître, l'équipe **attaquante** doit avancer d'une zone à l'autre avec tous ses joueurs, sans perdre le ballon pour marquer un essai.

Enfin, on trouve 10 occurrences de participe présent (ou gérondif), toutes suivies du complément du verbe (dont 4 occurrences pour le verbe pronominal *s'attaquant*) :

22) Celles-ci représentent le plus souvent les visages sérieux et fervents d'activistes qui ont sacrifié leur vie pour la cause palestinienne au nom d'Allah, souvent **en attaquant** des colonies juives aujourd'hui désertes.

<sup>23</sup> Ces occurrences proviennent pour la plupart de comptes rendus sportifs.

Le cas de *correspondant* se présente tout à fait différemment, puisque c'est ici l'emploi adjectival qui domine largement (61,9% du total), particulièrement pour les formes au féminin :

23) Vous trouverez les descriptifs et les coordonnées dans les rubriques **correspondantes**.

L'emploi nominal, présent particulièrement dans les formes masculines, ne représente lui qu'un bon quart du total des occurrences :

24) Les rédactions montent des dossiers, fouillent leurs archives, consultent **leurs correspondants** du Nord au Sud, de l'Est à l'Ouest de notre bonne vieille planète.

Enfin, le participe présent invariable intervient dans 100 occurrences, toujours accompagné de son complément, sous la forme *correspondant à* :

25) Sont utilisés les sigles **correspondant à** la dénomination du produit.

On notera que la règle du participe présent invariable pose manifestement des problèmes aux scripteurs ; en effet, on trouve dans le corpus 13 occurrences de participes accordés (6 au masculin pluriel, 7 au féminin pluriel)<sup>24</sup>, comme dans :

26) Tapez votre mot-clé, le moteur recherche les pages **\*correspondantes à** votre demande dans tout le site.

### 4.3 Les formes en *-able*

Deux formes en *-able* ont été retenues pour ce travail : *contribuable* et *coupable*.

terme de recherche	contribuable	contribuables	total formes
nbre d'occurrences (nbre total)	200 (924)	200 (2 062)	400 (2 986)
nom	198	200	<b>398 (99,5%)</b>
adjectif	2	0	<b>2 (0,5%)</b>

terme de recherche	coupable	coupables	total formes
nbre d'occurrences (nbre total)	200 (5 331)	200 (1 799)	400 (7 130)
nom	29	70	<b>99 (24,8%)</b>
adjectif	171	130	<b>301 (75,2%)</b>

Tableau 5 : distribution de deux formes en *-able*

Le paradigme des formes en *-able*<sup>25</sup> ne compte que deux formes, chacune épïcène : l'une pour le singulier et l'autre pour le pluriel. *Contribuable* s'utilise presque exclusivement comme

<sup>24</sup> Dans la mesure où ces cas ne remplissent pas les critères morphologiques pour être considérés comme des participes présents, ils ont été classés parmi les adjectifs dans la présente étude.

nom (*le/la/les contribuable-s*), soit à 99,5%. Au singulier, une seule forme se trouve avec l'article féminin (27) ; les autres sont majoritairement des masculins à valeur générique (cf. 28 ; en l'absence de contexte, il n'est cependant pas toujours possible de distinguer entre un emploi générique et un emploi spécifique, c'est-à-dire pour faire référence à un homme contribuable particulier).

- 27) Pour Mon Repos, le raisonnement de **la contribuable** n'est pas pertinent.
- 28) Pendant que **le contribuable** risque de payer les pots cassés [...]

Parmi les 200 occurrences sélectionnées au singulier, seules deux se rapprochent d'un usage adjectival (29, 30) ; au pluriel, aucun usage adjectival ne se laisse observer :

- 29) Un comble pour le téléspectateur **contribuable** [...]
- 30) [...] le même public mais version **contribuable** est prié de payer.

Quant à *coupable*, ce mot s'utilise majoritairement comme adjectif, notamment dans des constructions de type (*être*) *reconnu coupable* ou *plaider coupable* (31, 32) :

- 31) En 1993, il fut reconnu **coupable** à tort d'un meurtre par balles.
- 32) Tous plaident non **coupables**.

Les formes nominales sont minoritaires, mais plus fréquemment utilisées au pluriel (*les coupables*) qu'au singulier (*le/la coupable*) (33, 34) :

- 33) Sans **coupable**, le dossier a été clos en avril 2002.
- 34) [I]l est normal qu'on veuille interpellier **les coupables**.

À remarquer que seules trois formes sur 29 (au singulier) et une sur 70 (au pluriel) s'utilisent avec un article (ou autre élément satellite) au féminin.

#### 4.4 Les formes en *-eur*

Les paradigmes des formes en *-(t)eur* (au masculin) ne se laissent pas décrire de manière simple, car les considérations de type morphologique ne sont pas les seules à être déterminantes, bien qu'elles soient souvent éclairantes : ainsi, on peut s'attendre à une forme féminine lorsqu'il existe un verbe correspondant à un masculin en *-eur* : *vendeur/vendeuse*, *travailleur/travailleuse*). Les formes en *-trice* (d'après le modèle latin) s'observent lorsqu'il existe des mots en *-ture* ou *-tion* (*créatrice*, *inspectrice*). Durant le XX<sup>e</sup> siècle, de nombreuses formes féminines en *-eure* ont vu le jour, qui sont formées selon le modèle *majeur/majeure* (cf. Coutier 2002 et Elmiger 2008: 99).

Nous avons sélectionné trois exemples de mots se terminant par *-(t)eur(e)/trice*, au singulier : *conservateur/conservatrice*, *créateur/créatrice* et *amateur-e/amatrice*.

---

<sup>25</sup> Pour une analyse des dérivés en *-able*, cf. Hathout et al. (2003), qui distingue entre différents modes de formation. La question de savoir si celui-ci a une incidence sur la distribution catégorielle ne peut pas être développée dans le présent article, mais mériterait assurément un examen plus approfondi.

terme de recherche	conservateur	conservatrice	conservateurs	conservatrices	total formes
nbre d'occurrences (nbre total)	200 (6 787)	200 (1 442)	200 (5 099)	200 (322)	800 (13 650)
nom	52	18	164	5	238 (29,8%)
adjectif	148	182	36	195	562 (70,3%)

terme de recherche	créateur	créatrice	créateurs	créatrices	total formes
nbre d'occurrences (nbre total)	200 (3 941)	200 (847)	200 (2 808)	200 (209)	800 (7 805)
nom	176	70	192	51	489 (61,1%)
adjectif	24	130	8	149	311 (38,9%)

terme de recherche	amateur	amateur	amatrice	amateurs	amateurs	amatrices	total formes
nbre d'occurrences (nbre total)	200 (3 274)	23	39	200 (6 871)	5	14	481 (10 226)
nom	127	10	34	169	0	10	350 (72,8%)
adjectif	73	13	5	31	5	4	131 (27,2%)

Tableau 6 : distribution de trois formes en *-(t)eur(e)/(t)euse/trice*

Seuls les mots-formes *conservateur-s* (m) et *conservatrice-s* (f) sont attestés ; la forme féminine *conservateur-s* n'apparaît pas dans CFL. *Conservateur* s'utilise majoritairement comme adjectif (70,3%) (35), sauf au masculin pluriel, où l'emploi nominal est majoritaire (*les conservateurs* désignant les personnes appartenant à courant (politique) conservateur, en général) (36).

35) [...] le chef de l'opposition **conservatrice**, le président du Parti populaire [...]

36) **Les conservateurs** restent donc sur les rails libéraux.

La catégorisation de certaines formes s'avère difficile : en l'absence d'indices contextuels, il n'est pas aisé de déterminer si *conservateur*, dans les exemples suivants, est à classer comme nom (dans le second exemple comme nom complément épithète de *chef*) ou comme adjectif :

37) **le conservateur** Art Hanger

38) le chef **conservateur** Stephen Harper

À notre avis, le premier (37) serait à considérer comme un nom, alors que le second (38) aurait une valeur adjectivale.



Les formes *créateur-s*, *créatrice-s* se sont révélées particulièrement difficiles à catégoriser du fait que dans de nombreux cas, aucun critère syntaxique n'a permis de décider si, dans les deux exemples suivants (de type X est *créatrice* de Y), *créatrice* a une valeur d'adjectif ou de nom :

39) la lutte contre l'inflation peut être **créatrice** de chômage (considérée comme adjectif)

40) Antoinette Vonder Mühl est elle-même **créatrice** de bijoux (considérée comme nom)

La catégorisation a ainsi dû se faire sur la base de critères sémantiques (sujets à variation), ce qui a pour effet que le décompte pour *créateur* (ainsi que, dans une moindre mesure, pour *amateur*), doit être pris avec un grain de sel.

Au masculin, *créateur-s* s'emploie majoritairement comme nom (à 92%) (41), tandis qu'au féminin, l'emploi adjectival est majoritaire (69,8%) (42) :

41) *Fluide Glacial* porte encore l'empreinte de **son créateur** et de ses premiers rédacteurs en chef.

42) Cette volonté **créatrice** en marche est forte.

La forme masculine *amateur-s* a deux formes féminines : *amateur-e-s* et *amatrice-s* ; les deux ont une fréquence relativement basse, mais comparable (les formes féminines sont attestées quelques douzaines de fois seulement, ce qui est peu par rapport aux quelque 10 000 occurrences des formes masculines). Au masculin, *amateur-s* est majoritairement utilisé comme nom (à 74%) (43), moins souvent comme adjectif (44). Dans les deux exemples, la frontière entre usage nominal et adjectival semble ténue (en 44, c'est surtout l'adjectif *professionnelle* qui suggère que *amateur* a une fonction d'adjectif), ce qui laisse présumer que dans le cas d'*amateur/amatrice*, un glissement catégoriel est en plein cours. D'ailleurs, seule une partie des dictionnaires considèrent ce mot comme nom *et* adjectif.

43) **Amateur** de voyages, ses pérégrinations l'ont [conduit ...].

44) [...] une joueuse de tennis néerlandaise, **amateur** dans les années 1960 puis professionnelle [...].

Au féminin, *amatrice-s* est plus fréquent que *amateur-e-s* ; *amatrice-s* a majoritairement une fonction nominale (83%) (45), tandis que la forme plus récente *amateur-e-s* semble plutôt s'utiliser comme adjectif (dans 18 cas sur 28) (46) ; néanmoins, le faible nombre d'occurrences ne permet pas de tirer des conclusions fiables.

45) La reine Mary, qui était **une** grande **amatrice** de miniatures.

46) Est-ce la culture **amateur** ou celle qui plaît au plus grand nombre ?

## 5 Conclusion

Dans ce travail, la proximité formelle et catégorielle entre adjectifs et noms a été étudiée à l'aide d'un grand corpus de français écrit. L'étude empirique permet de constater des disparités considérables non seulement entre les quatre cas de figure retenus, mais aussi entre les divers exemples analysés dans chacune d'entre elles. Il apparaît clairement qu'il n'est pas possible de généraliser les résultats, car la distribution catégorielle d'un nom susceptible d'être utilisé comme adjectif ou d'un adjectif pouvant être employé comme nom semble être propre à chaque cas individuel. À côté de mots où différents usages sont attestés – et codifiés dans les ouvrages de référence – depuis longtemps (p. ex. *correspondant* comme participe

présent, adjectif et nom), certains cas laissent présumer un glissement catégoriel en cours (p. ex. *amateur/amatrice*, dont la valeur adjectivale, très proche de la valeur nominale, ne semble pas encore communément admise).

Non seulement les résultats permettent d'obtenir un éclairage empirique sur les phénomènes théoriques de transcatégorisation, mais encore la démarche présentée fournit de plus amples connaissances lexicales exploitables en lexicologie et dans l'enseignement du français langue étrangère (FLE). Si la question de la transcatégorisation ne constitue pas forcément une priorité dans l'enseignement du FLE à un niveau débutant ou intermédiaire, elle peut en revanche s'avérer très intéressante pour des niveaux avancés (de type C1 ou C2). En effet, si les locutrices et locuteurs natifs d'une langue sont apparemment capables de reconnaître la variation catégorielle du lexique et de l'utiliser productivement dans le discours, cela n'est pas nécessairement le cas pour des personnes qui sont en train d'apprendre la langue. Dans les grammaires de référence, on peut trouver des descriptions de phénomènes comme la conversion, mais les catégories nominale et adjectivale sont en général traitées de manière distincte. Connaître le paradigme formel, le sens et la valence d'un verbe ne suffit pas, par exemple, pour savoir si ses participes peuvent s'utiliser comme un nom commun de personne (*abonné·e* oui, mais pourquoi pas *abonnant·e*, comme p. ex. en allemand : *Abonnent·in* ?). Ce type de connaissance semble particulièrement utile pour des personnes parlant comme L1 des langues où les phénomènes de transcatégorisation sont fréquents, mais n'obéissent pas nécessairement aux mêmes régularités qu'en français (p. ex. en allemand, où les noms déverbaux (provenant de participes) sont très fréquents, cf. Elmiger 2011).

Il nous semble en outre que la méthode utilisée (l'étude détaillée d'un échantillon aléatoire simple de 200 occurrences) ne doit pas obligatoirement se limiter à la recherche, mais pourrait aussi s'utiliser dans l'enseignement dans le cadre d'une démarche d'apprentissage sur corpus, pour permettre à l'apprenant·e de découvrir la variation sémantique d'une unité lexicale, les catégories grammaticales qu'elle peut endosser, les locutions fréquentes dans lesquelles elle apparaît, etc. Pour reprendre les catégories de Boulton et Tyne (2014: 7) présentées *supra*, on passerait ainsi d'une utilisation « indirecte » des corpus pour l'apprenant·e, telle qu'elle a été décrite ci-dessus, à une utilisation « directe » par l'apprenant·e. Pour des raisons de temps, il ne sera certes pas toujours possible de faire l'exercice *in extenso* (soit avec 200 occurrences pour chaque forme fléchie), mais un nombre plus restreint d'occurrences (p. ex. 50 pour une ou deux formes) peut déjà fournir des résultats instructifs, avec un effort raisonnable à fournir (tant au niveau de l'étudiant·e qu'à celui de la personne qui supervise le travail). Certes, cette méthode ne permet pas d'attester des phénomènes rares, mais ceci ne saurait guère être un désavantage dans le contexte du FLE, où l'intérêt premier devrait être une bonne maîtrise des usages fréquents du lexique, avant l'intégration de cas plus marginaux.

Le présent travail s'est fait dans une perspective monolingue ; néanmoins, il serait intéressant de le prolonger dans une vision plurilingue, à l'aide de corpus parallèles ou comparables (cf. Kamber/Dubois 2014), ou de corpus contenant des productions d'apprenant·e·s. De la sorte, les particularités du français (et les difficultés à les apprendre) pourraient être décrites plus en détail, que ce soit au niveau théorique, lexical ou didactique.

## Bibliographie

- Apothéloz, Denis (2002): « La construction du lexique français. Principes de morphologie dérivationnelle ». In: *Journal of French Language Studies* 14/1. Paris/Gap, Ophrys. doi: 10.1017/S0959269504221585.
- Blumenthal, Peter/Novakova, Iva/Siepmann, Dirk (eds.) (2014): *Les émotions dans le discours / Emotions in Discourse*. Frankfurt a. M.: Lang.
- Boulton, Alex/Tyne, Henry (2014): *Des documents authentiques aux corpus: démarches pour l'apprentissage des langues*. Paris: Didier.
- Coates, Jennifer (1983): *The semantics of the modal auxiliaries*. London/Canberra: Croom Helm.
- Coutier, Martine (2002): « Le féminin des noms de personne en -(t)eur: résistances et concurrences ». In: Mathieu, Marie-Jo (sous la direction de): *Extension du féminin. Les incertitudes de la langue*. Paris, Honoré Champion: 69–93.
- Dal, Georgette/Hathout, Nabil/Namer, Fiammetta (1999): « Construire un lexique dérivationnel: théorie et réalisations ». *Conférence TALN*. Cargèse. <http://stl.recherche.univ-lille3.fr/sitespersonnels/dal/1999TALN.pdf> [23.07.2016].
- Eckart, Thomas/Elmiger, Daniel/Kamber, Alain/Quasthoff, Uwe (2013): *Frequency Dictionary French/Dictionnaire de fréquence du français*. Leipzig: Leipziger Universitätsverlag.
- Elmiger, Daniel (2008): *La féminisation de la langue en français et en allemand. Querelle entre spécialistes et réception par le grand public*. Paris: Honoré Champion.
- Elmiger, Daniel (2011): „Von Dozierenden und Emeritierenden: substantivierte Partizip-I-Formen im heutigen Deutsch“. *Travaux Neuchâtelois de Linguistique* 55: 163–179.
- Elmiger, Daniel (2013): « Pourquoi le masculin à valeur générique est-il si tenace, en français ? ». *Romanica Olomucensia* 25/2: 113–119.
- Elmiger, Daniel (2015): *Les corpus Bundesblatt / Feuille fédérale / Foglio fédérale. V. 1.2*. Genève.
- Gougenheim, Georges (1958): *Dictionnaire fondamental de la langue française*. Didier: Paris.
- Hahn, Marion et al. (2008): „elexiko – das elektronische, lexikografisch-lexikologische korpusbasierte Wortschatzinformationssystem. Zur Neukonzeption, Erweiterung und Revision einzelner Angabebereiche“. In: Klosa, Annette (ed.): *Lexikografische Portale im Internet*. Mannheim, Institut für Deutsche Sprache: 57–85. (= OPAL - Online publizierte Arbeiten zur Linguistik 1/2008).
- Haspelmath, Martin/Sims, Andrea (2010): *Understanding Morphology*. London: Hodder Education.
- Hass, Ulrike (ed.) (2005): „Besonderheiten des Gebrauchs“. In: *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz*. Tübingen/Berlin, de Gruyter: 265–276.
- Hathout, Nabil (2001): « Analogies morpho-synonymiques. Une méthode d'acquisition automatique de liens morphologiques à partir d'un dictionnaire de synonymes ». TALN 2001. [https://atala.org/doc/actes\\_taln/AC\\_0041.pdf](https://atala.org/doc/actes_taln/AC_0041.pdf) [23.07.2016].
- Hathout, Nabil/Plénat, Marc/Tanguy, Ludovic (2003): « Enquête sur les dérivés en –able ». *Cahiers de Grammaire* 28: 49–90.

- Hathout, Nabil/Sajous, Franck/Tanguy, Ludovic (2009): "Looking for French deverbal nouns in an evolving Web (a short history of WAC)". *Fifth Workshop on Web As Corpus*. San-Sebastian: 37–44. <https://halshs.archives-ouvertes.fr/halshs-00414494/document> [23.07.2016].
- Kamber, Alain (2011): « Contexte et sens: utilisation d'un corpus écrit dans l'enseignement/apprentissage du FLE ». *Travaux Neuchâtelois de Linguistique* 55: 199–218.
- Kamber, Alain/Dubois, Maud (2014): « Io vado n'est pas je vais. *Aller*, un verbe « intransitif » pour les apprenants de FLE ». In: Avanzi, Mathieu et al. (eds.): *Enseignement du français: les apports de la recherche en linguistique. Réflexions en l'honneur de Marie-José Béguelin*. Bruxelles, Lang: 131–149. (= GRAMM-R. *Études de linguistique française* 21)
- Kamber, Alain (2014): « Prendre, un verbe support dans l'enseignement du FLE: une analyse sur corpus ». In: Evouna, Jacques/ Essono, Onguéné/Martin, Louis (eds.): *Au cœur du verbe. Discours, syntaxe et didactique. Revue Mosaïques* 2: 3–16.
- Kerleroux, Françoise (1996): *La coupure invisible*. Lille: Presses Universitaires du Septentrion.
- Kübler, Natalie (2014): « Mettre en œuvre la linguistique de corpus à l'université. Vers une compétence utile pour l'enseignement/apprentissage des langues ? ». *Recherches en didactique des langues et des cultures : Les Cahiers de l'Acedle* 11/1: 37–77.
- Lecolle, Michelle (2011): « Désadjectivaux formés par conversion et double catégorisation: les cas des adjectifs/noms en *-aire* ». *Revue Romane* 46/2. Benjamins: 295–316.
- Lecolle, Michelle (2012): « Néologie sémantique et néologie catégorielle: quelques propositions ». *Cahiers de lexicologie* 100. Didier: 81–104.
- Mel'čuk, Igor (1993): *Cours de morphologie générale. Volume 1: Introduction et Première partie : Le mot*. Montréal: Presses de l'Université de Montréal.
- Mel'čuk, Igor A. (1996): *Cours de morphologie générale. Troisième partie: Moyens morphologiques; quatrième partie: Syntactiques morphologiques*. Montréal: Presses de l'Université de Montréal.
- Noailly, Michèle (1990): *Le substantif épithète*. Paris: Presses universitaires de France.
- Quasthoff, Uwe et al. (2014): "High Quality Word Lists as a Resource for Multiple Purposes". *Conference on Language Resources and Evaluation*: 2816–2819.
- Siepmann, Dirk (2015): "Dictionaries and Spoken Language: A Corpus-Based Review of French Dictionaries". *International Journal of Lexicography* 28/2: 139–168.
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Steyer, Kathrin (2013): *Usuelle Wortverbindungen: Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr.
- ten Hacken, Pius/Panocová, Renáta (2013): „The Use of Corpora in Word Formation Research”. *Cognition, Representation, Language*. <http://corela.revues.org/3018> [23.07.2016].
- Tutin, Agnès/Grossmann, Francis (2003): *Les collocations: analyse et traitement*. Amsterdam: de Werelt.
- Williams, Geoffrey (2006): « La linguistique de corpus: Une affaire prépositionnelle », *Texte*: 151–158. [www.revue-texto.net/Parutions/Livres-E/Albi-2006/Williams.pdf](http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Williams.pdf) [23.07.2016].