



HAL
open science

Du métadiscours au dictionnaire

Inka Wissner

► **To cite this version:**

Inka Wissner. Du métadiscours au dictionnaire: Pour un étiquetage unifié des mises en relief en lexicographie francophone. Gaétane Dostie; Pascale Hadermann. La dia-variation en français actuel. Études sur corpus, approches croisées et ouvrages de référence, Peter Lang, pp.297-317, 2015. hal-03607316

HAL Id: hal-03607316

<https://hal.science/hal-03607316>

Submitted on 16 Mar 2022

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

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

Du métadiscours au dictionnaire

Pour un étiquetage unifié des mises en relief en lexicographie francophone

Inka Wissner
Institut des Langues Romanes, Université de Zurich

1. Introduction

Les avancées de la linguistique française encouragent fortement les linguistes et les lexicographes à exploiter de plus en plus de corpus pour la préparation de dictionnaires. Dans ce contexte, le choix des corpus à utiliser représente un enjeu majeur, tout comme les multiples problèmes qui sont liés à leur exploitation. Ces difficultés, au cœur de débats animés, ne doivent pourtant pas détourner l'attention de la façon dont les données rassemblées sont présentées au sein des articles de dictionnaire.

La lexicographie française dispose d'excellents modèles microstructurels, constamment améliorés. Mais comment la microstructure rend-elle compte de la richesse des informations que rencontrent les lexicographes dans les corpus ? C'est aussi à la métalexigraphie qu'il revient de décrire le rôle et l'apport de ces ensembles textuels. Leur exploitation attentive révèle des renseignements fondamentaux sur les unités lexicales, mais aussi sur la nature foncièrement réflexive de la langue : le lexicographe expérimenté a maintes fois rencontré des passages où un énonciateur 'commente' plus ou moins indirectement la langue. Ces passages sont exploités depuis belle lurette en philologie et en linguistique françaises, et ont été commentés ici et là par les lexicographes, comme l'illustrent déjà des dictionnaires de l'ancien régime. Le présent article s'intéresse à la prise en compte de tels passages dans la microstructure d'articles de dictionnaires du français moderne. Après un bilan critique de la situation actuelle en lexicographie française (sections 2.1 et 2.2), il propose un modèle lexicographique, conçu pour la lexicographie différentielle francophone (sections 2.3 et 2.4). Il analyse ensuite deux phénomènes marqués du français contemporain pour illustrer l'étiquetage de leurs usages 'commentés' dans des articles de dictionnaire (section 3). Il s'appuie sur un dépouillement des plus grands ensembles textuels francophones qui permettent des requêtes lexicales et morphosyntaxiques, et qui sont rendus disponibles à la communauté scientifique. Par manque d'espace, il ne peut toutefois pas s'agir d'explicitier leur choix et les difficultés de leur exploitation dans une visée diatopique (voir Wissner 2012).

L'analyse proposée implique une série de positions scientifiques qui ne seront pas discutées dans le cadre de cet article. Elle présuppose :

- primo que la variation est une dimension inhérente à la langue (Coseriu 1981 : 12) ;
- secundo que le *métadiscours* fait partie intégrante du discours, tout comme le *métalangage* fait partie intégrante de la langue ;
- tertio qu'un dictionnaire de langue doit tenir compte du métadiscours, puisque ce dernier constitue, avec la *métalangue*, le métalangage (cf. Rey-Debove 1997 : 20sq. pour ces termes).

2. L'étiquetage des énoncés autoréférentiels en lexicographie française

2.1 Etat des lieux

Dans la lexicographie française, les exemples qui illustrent des énoncés métalinguistiques (ou *épilinguistiques*, suite à Culioli), dits *autoréférentiels* (c'est-à-dire qui renvoient à eux-mêmes, 2.2), du type *comme on dit* ou *X, que l'on appelle ainsi parce que*, ont bien entendu toute leur place. On en relève autant dans des dictionnaires 'généraux' que dans des dictionnaires spécialisés qui portent sur des aspects particuliers de la langue. Or, à l'heure actuelle, les seuls dictionnaires à avoir distingué et autonomisé de tels exemples dans la microstructure sont des dictionnaires de diatopismes qui relèvent de la lignée des travaux philologiques à visée historique, depuis le *Dictionnaire suisse romand* (Thibault 1997 : DSR).

Certains dictionnaires de diatopismes n'ont pas retenu ce balisage, comme le projet québécois, qui arrivait en 1998 au terme de très longues années de préparation (DHFQ). Il en est de même de dictionnaires plus récents comme le *Dictionnaire des belgicismes* (Francard et al. 2010 : DictBelg) et le *Dictionary of Louisiana French* (Valdman et al. 2010 : DLF), qui s'appuient essentiellement sur des enquêtes orales.

L'étiquetage a toutefois bien été appliqué par les contributeurs au *Dictionnaire des régionalismes de France* (Rézeau 2001 : DRF), qui en a augmenté la systématisme (cf. Thibault 2005, 96sq. et 102 ; Rézeau 2004, 122). Cette approche a ensuite été prolongée par l'éditeur, P. Rézeau (DRF Compl 2007, DRFA 2007). L'exemple est aussi suivi par différents lexicographes publiant des articles sur le français dans diverses aires de la francophonie dans deux volumes intitulés *Richesses du français et géographie linguistique* (RichessesRéz 2007, RichessesThib 2008).

Comment se présente la mise en valeur des citations autonomisées qui illustrent des énoncés autoréférentiels ? Dans les ouvrages cités, de telles citations figurent avant les exemples ordinaires, et sont précédées d'un alinéa qui étiquette le type d'énoncé autoréférentiel. Les étiquettes proposées dans le DSR et le DRF ont déjà été synthétisées dans le classement d'« exemples métalinguistiques » d'A. Thibault. Sous l'hyperonyme « en emploi métalinguistique » ont ainsi été regroupées cinq catégories principales : « en emploi métalinguistique », « en emploi métalinguistique ou autonymique » (c'est-à-dire autonymie avec un discours métalinguistique), « en emploi autonymique » (c'est-à-dire sans commentaire métalinguistique explicite), « avec un commentaire métalinguistique incident » (qui entoure ou précède un diatopisme) et « en énoncé définitionnel syntaxiquement autonome » (Thibault 2005 : 96-115, surtout 103-105). On trouve enfin également dans ces dictionnaires des balises rares comme « en apposition » (p. ex. DRF *cébette, clouque*) ; celle-ci concerne des définitions par simple juxtaposition.

Les recueils d'articles ultérieurs proposent un étiquetage similaire des 'exemples métalinguistiques'. A titre de comparaison, nous présentons ici les choix de trois contributeurs du recueil des RichessesRéz (ci-dessus) : Chambon, Chauveau et Rézeau. Dans leurs articles, ceux-ci ont fait précéder les citations qui illustrent des diatopismes de huit étiquettes différentes. Trois d'entre-elles (1, 3 et 5) correspondent à celles du DSR et du DRF :

- 1) « en emploi autonymique », p. ex. *X qu'on appelle ici Y* ;
- 2) « dans un énoncé définitoire ordinaire », p. ex. *un Y, c'est X* ;
- 3) « avec un commentaire métalinguistique incident », p. ex. *les X, c'est-à-dire Y* (RichessesRéz 2007 : 208-253 [Chauveau], 272-305 [Rézeau], aussi pour les exemples) ;
- 4) « dans un commentaire métalinguistique incident », p. ex. *X qu'on nomme Y, peut-être parce que ...* (v. 3) ;
- 5) « en emploi métalinguistique » (p. ex. *il dit maintenant Y*) ;
- 6) « suivi d'un commentaire métalinguistique », p. ex. *X, c'est-à-dire Y* (catégorie plus proche d'une définition que d'un commentaire comme sous 4) (RichessesRéz 2007 : 208-253 [Chauveau] aussi pour les exemples).

S'y ajoutent d'autres catégories plus rares dans les articles de J.-P. Chambon, comme les suivantes :

- 7) « avec marque de connotation autonymique » (RichessesRéz 2007 : 263, 264, 268, 269 p. ex.) (au sujet de marquages typographiques) ;
- 8) « en emploi autonymique, sans marque » (RichessesRéz 2007, 342 p. ex.) (il s'agit de gloses sans marquage typographique).

Sans retracer ici l'histoire complexe et instructive de la mise en place des projets lexicographiques cités, nous nous concentrerons sur la forme finale des ressources lexicographiques sélectionnées.

2.2 Analyse critique

Les ressources lexicographiques analysées présentent une avancée importante en lexicographie française : elles offrent une meilleure prise en compte que les sources antérieures de la diversité des manifestations métadiscursives de la langue qu'attestent les corpus. Elles affichent toutefois quelques aspects qui gagneraient à être revus. Ainsi, certains exemples qui illustrent des diatopismes typographiquement marqués (par exemple à l'aide d'italiques) n'ont pas été autonomisés dans la microstructure du DSR (cf. Thibault 2005, 105sq.) et du DRF (2001). En outre, certaines étiquettes

proposées sont incompatibles, non seulement entre lexicographes (problème de coordination), mais aussi parfois chez un même auteur (p. ex. RichessesRéz 2007 : 435, 451sq., 466 p. ex. [Rézeau] et 341-343, p. ex. [Chambon/Grémois]).

Il faut savoir que les étiquettes ont été élaborées de façon plutôt intuitive, selon les besoins qui se présentaient au fil de l'avancement des travaux (comm. pers. de J.-P. Chambon, 2010). Dans le cadre de la préparation de dictionnaires comme le DSR et le DRF, leur attribution était loin de constituer un objectif majeur. Vu la nature foncièrement pragmatique de l'élaboration de dictionnaires innovants, un certain réalisme s'impose : une fois un projet lancé, le lexicographe n'a pas toujours le loisir de revenir à l'analyse comparative de travaux antérieurs comme nous pouvons le faire ici (sections 2.1 et 2.3), ni de formaliser un nouveau modèle descriptif (section 2.4).

Lors du choix de dénomination des étiquettes, les ouvrages analysés ont bien intégré des concepts fondamentaux de la métalexigraphie, considérablement marquée par les travaux de Rey-Debove, comme *métalangage*, *autonymie* et *connotation*. En même temps, l'intégration de la terminologie rey-debovienne n'a pas été systématique et s'avère, dans certains cas, inexacte. Ainsi, des diatopismes qui sont présentés comme étant « en emploi autonymique » (DRF, RichessesRéz) figurent presque exclusivement en *connotation* autonymique. En effet, ils ne renvoient pas simplement au mot lui-même (comme dans *le mot de X : emploi autonymique*) ; ils connotent aussi le signe tout en ayant une fonction référentielle (comme dans *X qu'on appelle aussi Y*). On parle alors dans de pareils cas, à la suite de Rey-Debove, de *connotation autonymique*. La distinction entre des *emplois* et des *connotations* autonymiques est donc essentielle pour décrire le fonctionnement des énoncés (section 2.3). Dans l'ensemble, on constate donc qu'une terminologie unifiée manque à l'heure actuelle en lexicographie différentielle francophone (Thibault 2005 : 103). Ce manque appelle à l'établissement d'un système d'étiquetage argumenté et unifié.

2.3 Apport d'autres disciplines à la lexicographie

En nous appuyant sur le système d'étiquetage actuel des 'exemples métalinguistiques' en lexicographie différentielle francophone, nous proposons d'intégrer les avancées de trois disciplines dont la confrontation permet d'élaborer une catégorisation descriptive unifiée. Nous avons exploité les travaux métalexigraphiques et sémiotiques de J. Rey-Debove (notamment 1997 [1978]). Ceux-ci ont été complétés des avancées en pragmatique, plus spécifiquement des considérations métaénonciatives de J. Authier-Revuz (1995) et des acquis en analyse du discours (Maingueneau 2005, 2009).

C'est en effet la sémioticienne Rey-Debove qui a popularisé en linguistique française une notion créée par le positiviste R. Carnap. Elle insistait d'abord sur la distinction entre des emplois 'ordinaires' (c'est-à-dire qui sont utilisés comme signifiants pour référer à un signifié) et des emplois autoréférentiels, ou 'signes autonymiques', c'est-à-dire qui renvoient à eux-mêmes en tant que mots (Rey-Debove 1997 : 57-164). Parmi ces derniers, elle distinguait en outre de simples emplois autonymiques (du type *As-tu déjà entendu le mot de X ?*) et des énoncés en 'connotation autonymique' (Rey-Debove 1997 : 260). Ils attirent l'attention sur un signe en tant que mot tout en ayant une fonction référentielle (comme dans *les X, comme ils les appellent*). Il s'agit des cas où un signe renvoie à lui-même, mais « recèle » en même temps « un effet de sens » (Rey-Debove 1997 : 260).

Le concept de « connotation » a été complexifié par celui de « modalisation autonymique » (Authier-Revuz 1995). Ce dernier intègre le premier, tout en l'élargissant à des énoncés qui véhiculent des attitudes concernant différents paramètres de la situation communicationnelle (en dehors du discours). Un énoncé modalisé peut en effet véhiculer l'attitude que porte l'énonciateur sur le discours même (Authier-Revuz parle d'une *non-coïncidence* « du discours à lui-même »). Dans ce cas, l'énonciateur fait allusion à un discours d'individus ou d'un collectif qui est en train de se tenir, ou à un discours antérieur (comme dans *elle souffrait d'avoir souvent été appelée X, terme vulgaire qui ...*). Bien évidemment, le signe est nécessairement utilisé par l'énonciateur ; or, lorsqu'un énoncé autoréférentiel modalisé porte sur le discours même, l'énonciateur déclare implicitement à la fois qu'il utilise le signe et qu'il le cite, en l'attribuant à autrui (cf. Authier-Revuz 1995 : 40). Il s'agit du type de connotation autonymique dont rend aussi compte J. Rey-Debove lorsqu'elle parle de différents *modes du dire* (qui peuvent être véhiculés obliquement ou directement), du type « comme il/elle dit » et « comme on dit » (1997 : 267-291). De tels concepts sont particulièrement pertinents pour des énoncés qui intègrent des diatopismes, ceux-ci étant souvent attribués à des (groupes de) locuteurs différents de ceux qui les ont produits. Ainsi la catégorie des *commentaires métalinguistiques* (section 2.1), portant sur le discours

d'autrui, est la plus fréquente dans le DRF (Thibault 2005 : 103), tout comme dans RichessesRéz (section 2.2).

Comme le précise J. Authier-Revuz, un énoncé en modalisation autonymique peut toutefois aussi véhiculer un écart « entre interlocuteurs » (signalant une différence d'usage entre les coénonciateurs), « entre mots et choses » (remettant en cause l'adéquation d'un emploi pour renvoyer au signifié), et « des mots à eux-mêmes » (lorsqu'il y a ambiguïté sémantique) (1995 : 100). Les modalisations autonymiques (ou 'prises en mention') sont signalées par des *modalisateurs*, comme des *mots métalinguistiques* du type *s'appeler* (Rey-Debove 1997 : 283), ou des signes qui véhiculent un jugement, tels que des adjectifs ou des verbes de citation (*s'écrier*, *insulter*, etc.). Ce sont là des éléments « grâce auxquels l'énonciateur peut dans le fil de son discours commenter sa propre parole » (Maingueneau 2005 : 177).

L'autonymisation des énoncés est à distinguer, dans le discours, des attributions aux unités discursives, qui relèvent donc de l'organisation textuelle. En suivant la terminologie proposée en analyse du discours, on peut distinguer trois grandes catégories de discours rapporté. Celle du *discours citant* concerne le discours qui est présenté comme assumé pleinement par l'énonciateur (dans le présent passage, par exemple, nous assumons pleinement le contenu et la forme des énoncés ici mis entre parenthèses). Une deuxième catégorie est celle des *discours cités*, c'est-à-dire des discours qui sont *présentés* comme cités, donc rapportés, que ce soit de façon explicite – on parle généralement de *discours direct* (comme dans *il s'est écrié : « j'en veux aussi ! »*) – ou de façon indirecte. Dans ce second cas, la source énonciative est ambiguë (Maingueneau 2005 : 116). Ainsi, dans un énoncé comme *Elle a assuré qu'elle n'en souhaitait point*, le contenu du discours cité est attribué à une autre instance énonciative (elle) ; en l'absence d'indices métadiscursifs explicites (italiques, gloses, etc.), il n'est toutefois pas clair si les choix lexicaux (*souhaiter*, *point*) et le recours à l'élément de négation *ne* sont dus à la locutrice citée, ou à l'énonciateur qui la cite. On distingue alors traditionnellement le discours indirect (comme dans l'exemple cité), le discours direct libre – discours rapporté qui a les propriétés linguistiques du discours direct sans toutefois être signalé comme rapporté à l'aide d'une signalisation (typographie, tournure de citation) (Maingueneau 2005 : 125) –, et le discours indirect libre. Il s'agit là également d'un discours rapporté non signalé comme tel, mais qui a les propriétés linguistiques du discours indirect (concordance des temps, etc.) (Maingueneau 2005 : 131). Dans ces types de discours 'hybrides', c'est-à-dire où se superposent les propos de deux ou plusieurs locuteurs, s'expriment plusieurs 'voix' : celle de l'énonciateur qui cite, et celle d'un (ou de plusieurs) locuteur(s) cité(s) (Maingueneau 2005 : 127-133). D'autres énoncés enfin peuvent être attribués à une troisième catégorie de discours : celle du *paratexte*. Par ce terme originellement proposé par Genette (1987), on entend l'ensemble des fragments verbaux qui accompagnent un texte proprement dit, qu'il s'agisse d'unités vastes (préfaces, texte de couverture, ...), ou d'unités réduites (titre, nom de rubrique, notes de bas de page, ...) (v. Maingueneau 2009 : 94).

L'apparition d'énoncés dans les unités du discours est significative lorsqu'il s'agit d'*attributions* discursives : elle facilite l'interprétation sémiotique, comme celle des mises en relief proprement dites. Or, elle peut aussi représenter le résultat d'une simple *distribution*. Contrairement aux distributions discursives, les emplois et modalisations autonymiques attirent toujours l'attention sur l'unité qu'ils commentent. Il importe toutefois de préciser que l'emphase peut être plus ou moins forte, selon la position des unités dans la phrase. Une unité en fonction de thème (comme dans *X, autrement dit ...*) renvoie à elle-même mais assume aussi une fonction référentielle dominante (il y a connotation ou modalisation autonymique). Dans un dispositif métadiscursif où une unité figure en fonction de rhème, comme dans *on l'appelle X*, l'énoncé attire davantage l'attention sur le mot même : l'unité est mise en valeur par une emphase forte et est essentiellement autoréférentielle.

Les concepts présentés ci-dessus n'ont pas été directement conçus pour l'analyse de réalisations discursives de diatopismes. J. Rey-Debove tient bien compte de la dimension diatopique de la langue, en considérant toutefois qu'elle relève de langages « étrangers » (1997 : 281). Quant à elle, J. Authier-Revuz parle au sujet du français de variation « dialectale » (1995 : 239-241), en se référant sans doute elle aussi à la variation diatopique du français même. Chez les deux auteures se perpétue ainsi une conception ancienne selon laquelle la variation serait extérieure à un système invariable. La linguistique variationnelle a toutefois largement démontré l'ancrage profond et le caractère intrinsèque de la variation (diatopique, diastratique, etc.) dans la langue française elle-même (section 1). Ainsi, les travaux métalexigraphiques, sémiotiques et pragmatiques que nous avons cités gagnent à intégrer les acquis de la linguistique variationnelle, tout comme la lexicographie peut s'enrichir grâce aux avancées des premiers.

2.4 Pour un nouvel étiquetage des citations en lexicographie différentielle

Les travaux sémiotiques et pragmatiques présentés ci-dessus proposent des distinctions conceptuelles qui permettent de mieux cerner la forme et le sens des différentes réalisations métadiscursives que le linguiste rencontre lors du dépouillement de corpus. Pour appliquer ces distinctions à un modèle d'étiquetage formel unifié pour la lexicographie différentielle francophone, il est pertinent de transférer le centre d'intérêt des unités autonymisées elles-mêmes, vers les dispositifs formels qui les présentent en emploi ou en modalisation autonymiques. Ce sont ces dispositifs, à la fois métadiscursifs et métalinguistiques, qui permettent une interprétation sémiotique du discours. Ce sont eux aussi qui attirent l'attention des utilisateurs des dictionnaires.

Pour désigner les différents types de dispositifs qui autonymisent les énoncés autoréférentiels (et non pas les énoncés eux-mêmes), nous avons choisi une étiquette nouvelle. Il s'agit du terme de *mise en relief*, pris au sens d'« emphase », dont la réalisation discursive résulte du « fait d'attirer particulièrement l'attention sur un des éléments de la phrase » (Grevisse 2008 : 575, §455). Ce terme est véritablement hyperonymique (contrairement à l'étiquette actuelle, 'en emploi métalinguistique', section 2.1) et permet de mieux rendre compte de l'effet des dispositifs désignés sur les coénonciateurs. Il vise en outre à diminuer les risques d'ambiguïté terminologique (présentés par l'étiquette actuelle), étant donné qu'il est explicitement placé dans un nouveau cadre descriptif. Il nous semble enfin davantage transparent pour les utilisateurs de dictionnaires.

La confrontation des catégories utilisées jusqu'à présent en lexicographie différentielle et l'intégration des avancées en métalexigraphie, sémiotique et pragmatique (section 2.3), permettent de distinguer deux grands groupes de *mises en relief* de diatopismes identifiés ci-dessous.

1) Les **marques (signes typographiques)**, c'est-à-dire :

- des mises entre guillemets, éventuellement réalisées à l'aide de *guillemets de citation*, dont les derniers peuvent signaler des îlots textuels (sur ce terme, v. Authier-Revuz 1995 : 134 *sq.*, 512-515), comme dans *Ils disent* « X » ;
- des mises en italiques ;
- d'autres éléments typographiques tels que des parenthèses, des tirets (doubles tirets, tirets de figement), des points de suspension ou encore l'impression en lettres différentes du texte principal, (lettres en caractères gras, en capitales, etc.).

Ces marques peuvent concerner chacun des paramètres décrits par Authier-Revuz (2.3). Dans le domaine de la diatopie, elles portent surtout sur l'usage de groupe de locuteurs.

2) Les **gloses**, manifestées sous l'une ou l'autre des formes suivantes :

- **les commentaires métalinguistiques incidents**, signalés par des modalisateurs du type *mot, dire* ou *appeler*. Il peut s'agir a) de simples *tournures autoréférentielles* du type *le mot de Y*, b) de *commentaires explicatifs* du genre *ils disent Y parce que* ou c) de *commentaires de citation* (p. ex. *ils l'appelaient Y, X que l'on appelait Y, Y comme on l'appelle chez nous*). Cette dernière catégorie inclut aussi des *tournures de citation localisatrices* qui introduisent un diatopisme dans le discours cité (direct ou indirect, parfois en îlot textuel), tout en l'attribuant à l'usage linguistique d'une région donnée – qu'il soit présenté comme collectif ou individuel (p. ex. *les habitants disent « Y »*) ;
- **les définitions**. On trouvera ainsi a) des *définitions ordinaires* du type *X, c'est (ici) Y* (de nature lexicographique ou paraphrastique), qui sont donc syntaxiquement autonomes et b) des mises en équivalence par *juxtaposition de binômes*, qui sont surtout synonymiques et hyperonymiques. Elles sont manifestées par la ponctuation (comme des virgules), la conjonction *ou*, des parenthèses, ou encore un ajout en bas de page.

Cette classification des mises en relief de diatopismes est à distinguer de la *distribution* des unités de la langue dans l'organisation textuelle :

- le *discours citant* ;
- les *discours cités* : *discours direct* vs *discours à source énonciative ambiguë* ;
- le *paratexte* (section 2.3).

Notre catégorisation a pour mérite d'être systématique et appuyée sur les avancées de disciplines incontournables en la matière ; ses catégories ne sauraient toutefois être définitives, ni étanches entre

elles. Si le linguiste aime catégoriser la langue, tout système d'étiquetage rencontre nécessairement des limites face à la diversité des réalisations discursives concrètes. Ces limites sont particulièrement bien illustrées par la catégorie des *tournures de citation localisatrices* : elle est proche à la fois d'un discours cité du type *X dit que Y* (distribution discursive) et des mises en relief classiques du type *X, originaire de Z, dit plutôt « Y »*. Cette sous-catégorie des *commentaires de citation* est spécifique de la mise en relief de diatopismes. Lorsque l'on s'intéresse à d'autres dimensions variationnelles collectives, les corpus révèlent aussi, bien entendu, des passages équivalents qui modalisent le discours de milieux sociaux (variation diaphasique – du type *les bourgeois disent « Y »*), par exemple, ou d'époques spécifiques (variation diachronique – du type *avant on appelait cela un Y*).

3. L'étiquetage formel des citations : études de cas

Pour exemplifier notre système d'étiquetage, nous examinons ci-dessous des extraits de corpus relatifs à deux phénomènes linguistiques marqués, à savoir l'usage des substantifs *batture* et *argent* au féminin.

3.1 Le diatopisme morphologique *argent* (substantif féminin)

Lorsque l'on consulte les récents dictionnaires généraux du français, on apprend que le substantif *argent*, désignant la monnaie métallique ou papier-monnaie, est de genre masculin (p. ex. Rob, PR, Usito) ; ceci correspond en effet à l'usage soigné en français de référence. Seul le TLF signale aussi l'emploi du substantif au féminin. Il l'accompagne d'une marque diastratique, « populaire » (en interprétant un passage littéraire chez un écrivain normand, Queneau 1933 [Frantext]), et d'une marque diatopique, « Canada ».

Nous avons ensuite dépouillé les grands corpus en ciblant particulièrement des séquences qui révèlent le genre de *argent*, comme des cooccurrences avec des adjectifs antéposés (*bon, beau, gros*, etc. au féminin, singulier et pluriel). L'accord de participes et d'adjectifs postposés ou la reprise du substantif à l'aide d'un pronom ont eux aussi été des indices. Les corpus du français des XX/XXI siècles présentent en effet *argent* au féminin en Amérique du Nord (FLI, ILQ, Hansard, I-FR), en particulier en cooccurrence avec *grosse* et *belle* (FLI). Le corpus web I-FR atteste quant à lui la séquence *grosse argent* jusqu'à nos jours [2013]. Si la recherche n'a pas permis d'identifier *argent* au féminin dans la BDTS ni dans le CFPQ, ceci n'est pas à considérer comme étant significatif. De fait, dans la plupart des cas, le genre de *argent* est 'neutralisé' par des séquences qui ne le révèlent pas.

On rencontre également le genre féminin dans des corpus en Europe francophone septentrionale, plus précisément occidentale. Contrairement aux corpus nord-américains, les cotextes identifiés sont rares, et révèlent un usage très familier (voire populaire). Ils sont localisés en Normandie (*belle argent*, Queneau 1933 [Frantext, ci-dessus]) et dans l'Ouest de la France (ESLO1 [Orléans, 1970 ; une attestation de 2010 est incertaine]).

L'usage de *argent* au féminin est confirmé par de rares dictionnaires spécialisés qui portent sur la variation diatopique, en Amérique du Nord (Poirier 1993 [1927-1933] [Acadie], DHFQ [Québec]) ainsi qu'en Europe. C'est le dictionnaire historique du FEW qui y confirme son emploi ancien (dès l'ancien français) ainsi que son emploi contemporain dans une large aire septentrionale (Belgique, France septentrionale, Suisse) (« pop. », « région. », dialectal, FEW XXV/145 : 192b, 195b ARGENTUM).

Parmi les nombreux exemples d'usage de *argent* s. f. que fournissent les grands corpus, quelques-uns seulement ont été mis en relief par les énonciateurs. Nous avons ici choisi de retenir les exemples contemporains auxquels donne accès Hansard, un immense corpus de débats parlementaires du Canada :

ARGENT s. f. "monnaie métallique ou papier-monnaie" (*grosse argent*). Réf. / Syn. gén. *argent* s. m.

Dans le discours cité (source énonciative ambiguë), avec mise en relief (commentaire de citation) :

[Québec/Ottawa] « Le principal problème au Québec est le sous-financement des universités et du réseau collégial. Au Québec, il y a les universités et les cégeps. Avec l'argent de la Fondation du millénaire qui est destiné aux étudiants, on croit que si on a un droit de retrait, cet argent devrait être réinvesti dans le régime d'aide financière [...]. J'aimerais souligner aux membres du comité des finances que le coût d'administration

de l'unité autonome d'aide financière au Québec, et cela ne répond pas à une question, est de 1,25 p. 100 [...]. Tout le monde dit que ce sera 5 p. 100. C'est **de la grosse argent, comme on dit chez nous**, qui ne sera pas **donnée** en bourse aux étudiants. » (Hansard [1998, témoignage du président de la Fédération étudiante universitaire du Québec prononcé devant le comité sénatorial permanent des finances nationales)

Dans le discours cité (source énonciative ambiguë), sans mise en relief :

[**Nouveau-Brunswick** : comté de Madawaska] « On a deux sources de revenu : l'aide du gouvernement pour nos finances, et la communauté. [...] En région urbaine, ils ont une capacité financière parce qu'il a des business qui font **de la grosse argent**. Les maisons mères, bien souvent, sont dans les centres urbains. » (Hansard [2007], témoignage du directeur général de l'Atelier RADO [Edmunston] prononcé devant le Comité sénatorial permanent de l'agriculture et des forêts)

a) Remarque métalinguistique

Dans l'exemple mis en relief, le type lexical est suivi d'un commentaire métalinguistique de citation. Celui-ci renseigne sur l'usage (véhiculant le mode du dire *comme on dit*), ainsi que sur la localisation de cet usage (*chez nous*, c'est-à-dire « au Québec »). La mise en relief indique que l'énonciateur est conscient que la séquence *grosse argent* n'est pas nécessairement partagée par tous les francophones et ne correspond pas à la norme (au sens de 'bon usage'), qu'il dit implicitement connaître. Cette conscience lui permet de justifier cette déviation par rapport au bon usage. En même temps, l'énonciateur dit indirectement que la séquence est pour lui transparente (il ne fournit pas de définition), et qu'il la juge légitime, même dans une situation communicative formelle comme un débat parlementaire. Elle relève à ce titre pour l'énonciateur de l'usage neutre (norme endogène).

b) Remarque formelle

L'usage du substantif au féminin représente dans le premier exemple une variante graphique (v. « argent [...] destiné [...] cet argent [...] »). Nous avons toutefois affaire à du discours hybride, puisque les deux passages cités ont été fixés à l'écrit par l'équipe du parlement canadien qui est chargé de la transcription des discours prononcés au parlement : la variation n'est donc pas à attribuer au locuteur québécois.

3.2 Le diatopisme sémantique *batture* (substantif féminin)

Le deuxième emploi marqué dont nous illustrerons l'étiquetage formel relève d'un type lexical qui est surtout associé au français québécois. Celui-ci connaît avant tout *batture* s. f. pour désigner un estran large à faible pente (typiquement au Saint Laurent) (Usito ; même « région. (Canada) » PR). Les grands corpus francophones donnent toutefois aussi accès à d'autres emplois. L'exemple retenu ici (formé sans doute à partir d'un sens "ce qui reçoit un choc"), est attesté dans l'usage ancien du grand Ouest en France (XVI-XVIII^e siècles) (Frantext), puis au Québec dans l'usage contemporain (I-FR) :

BATTURE s. f. « techn. mar. » "fond de roche ou de corail situé un peu au-dessous de la surface de l'eau, sur lequel la mer se brise" (sans syn. gén.)

Dans le discours citant, avec mise en relief (définition) :

[en référence au Québec (2002)] « schorre adjacent à la terrasse Mitis ou à l'âge de la **batture (faciès de la slikke)** sous le schorre supérieur » (I-FR. Source : Dionne, 2002, « Une nouvelle courbe du niveau marin relatif pour la région de Rivière-du-Loup (Québec) », *GPQ* 56/1, article scientifique publié en ligne sur www.erudit.org)

Dans le discours citant, sans mise en relief :

[*idem*] : « changement de profil longitudinal de la vasière (slikke ou **batture**) » (I-FR : même source)

Que remarque-t-on sur le plan métalinguistique ?

Dans l'exemple mis en relief, *batture* est suivi d'une définition à l'aide d'un syntagme qui vise à rendre le mot transparent. La présence de la mise en relief montre que l'énonciateur est soucieux d'être clair pour ses interlocuteurs (lecteurs d'un article scientifique). Vu la nature du document cité et

¹ Dans les exemples, la marque (mise en gras) a été ajoutée par nous-même pour faciliter la lecture des citations.

vu le contenu de la mise en relief – le syntagme comportant des termes techniques² – le diatopisme est lui aussi considéré par l'énonciateur comme un terme technique. La localisation du référé, au Québec, n'est quant à elle pas fournie dans la mise en relief, et seulement dans le titre du document, accessible par une recherche sur le web.

Dans le second exemple, le diatopisme n'est pas mis en relief lui-même. Il figure pourtant au sein d'une mise en relief, plus précisément une définition. Celle-ci se présente sous la forme d'une apposition d'hyponymes entre parenthèses : deux technicisms qui visent à rendre transparent un autre terme technique (général) (*vasière*). Il ne s'agit toutefois pas d'une définition ordinaire, exacte, l'apposition servant à spécifier deux types de référés particuliers auxquels peut renvoyer *vasière*. Le cotexte signale que les deux hyponymes sont considérés comme accessibles pour les interlocuteurs visés (spécialistes du domaine). Le corpus exploité illustre qu'il s'agit avec *batture* d'un usage qui est marqué à la fois diatopiquement et diaphasiquement.

Du point de vue technique, on remarquera que le type d'analyse mené ci-dessus n'est pas de nature semi-automatisée, n'ayant pas été envisagé par les réalisateurs des grands corpus existants du français dans la francophonie. Ils ne proposent à ce titre pas d'annotation des unités de discours, ni des mises en relief. Ce type d'étiquetage, très lourd, n'est en effet fourni que dans certains corpus de petite taille. Pour la plupart, les grands corpus ne fournissent pas non plus une annotation de type sémantique et grammaticale fine. Une telle annotation permettrait pourtant aux utilisateurs de retracer facilement les caractéristiques des unités de langue recherchées – ici des sèmes différents (pour *batture* s. f.) et des différences de genre (pour *argent* s. f.). La recherche menée a donc impliqué d'analyser chaque attestation des unités retenues dans les différents corpus (souvent recherchées par forme, au singulier et au pluriel), pour identifier leur sens exact, mais aussi les caractéristiques grammaticales de *argent*. Pour ce dernier, c'est en particulier l'examen du comportement syntagmatique qui a apporté des réponses (section 3.1).

4. Conclusions

Sur notre chemin *des corpus aux ouvrages de référence*, nous nous sommes concentrée sur la dimension métadiscursive de l'usage de diatopismes dans les grands corpus du français moderne, en ciblant la manière dont ils peuvent être étiquetés en lexicographie différentielle.

Un rapide bilan de l'état de l'art a permis de montrer la nécessité de proposer un étiquetage systématique pour rendre compte à la fois des unités de discours dans lesquelles figurent les diatopismes et des types de dispositifs métalinguistiques formels qui les accompagnent (*mises en relief*). La proposition d'étiquetage s'appuie à la fois sur l'usage actuel dans les dictionnaires de diatopismes et sur une exploitation de corpus. Elle allie pour cela les approches de la lexicographie et de la métalexigraphie, de l'analyse de corpus ainsi que de la sémiotique et de courants de la pragmatique. Elle vise à mieux rendre compte de la langue dans ses réalisations concrètes, y compris de ses usages métadiscursifs – usages qui individuellement relèvent du *discours*, mais qui font partie intégrante du fonctionnement de la *langue*, dont doit tenir compte le dictionnaire.

L'étiquetage reste à être appliqué dans le cadre d'un grand projet dictionnaire. Nous avons dû nous contenter ici d'en illustrer l'apport à la lexicographie francophone à travers l'exemple de deux phénomènes marqués : l'un diatopiquement et diaphasiquement (*batture*) (section 3.1), l'autre à la fois diatopiquement, diaphasiquement et sans doute diastratiquement (*argent* s. f.) (section 3.2). Comme le montre leur analyse, la présence d'étiquettes qui attirent l'attention sur les mises en relief dans un article de dictionnaire permet de cerner aisément les comportements métadiscursifs des phénomènes linguistiques. Dans les corpus retenus (Hansard, I-FR), les mises en relief sont de plus décisives pour identifier leurs caractéristiques dénotatives et variationnelles. Dans le cas de *argent* s. f., ces dispositifs révèlent le genre, identifiable par son fonctionnement phraséologique (*grosse argent*), mais aussi la localisation du terme dans l'espace (au Québec) et son usage neutre. Dans le cas du substantif *batture*, les renseignements que fournit la mise en relief portent à la fois sur son sémantisme et sur son appartenance variétale (technicisme) (section 3.2). Les deux exemples viennent ainsi souligner

² Dans les exemples analysés, le diatopisme est accompagné de quatre termes techniques : *schorre* « géogr. », et *vasière* (TLF), ainsi que *faciès* « géol. » (qui désigne des roches ou dépôts de fossiles) et *slikke* « sciences de la terre » (qui désigne un fond vaseux) (Foucault/Raoult 2010).

l'importance des grands corpus pour analyser non seulement les caractéristiques formelles et sémantiques des unités du français moderne, mais aussi leurs dimensions variationnelles.

En restant cohérente avec notre approche rigoureusement interdisciplinaire, nous nous permettons de conclure en ouvrant le débat sur l'apport possible du modèle d'étiquetage ici proposé. Ne serait-il pas pertinent d'intégrer les acquis de la lexicographie différentielle dans d'autres domaines de la lexicographie française – en l'occurrence l'étiquetage systématique dans la microstructure des mises en relief des unités de la langue ? Le modèle développé ici pourrait en effet porter sur différentes dimensions variationnelles du français – diastratique, diaphasique ou diamésique – voire sur la langue 'générale'. Si certaines catégories identifiées seraient à adapter pour un tel transfert (section 2.4), celui-ci contribuerait sans doute à une meilleure représentation lexicographique des usages effectifs, ordinaires et métadiscursifs, que le lexicographe rencontre dans les corpus.

Références

- Authier-Revuz, J., 1995, *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, Paris : Larousse.
- BDTS. Banque de données textuelles de Sherbrooke recueillant plus de 52 millions de mots de divers types de textes des années 1960 à 2000, préparée sous la direction d'H. Cajolet-Laganière et P. Martel, Sherbrooke ; sous-ensemble d'environ deux millions d'occurrences en accès limité via un index lexical, <<http://catfran.flsh.usherbrooke.ca/catifq/bdts/index.htm>> (05/11/2013).
- CFPQ, 2006- , *Corpus de français parlé au Québec* préparé sous la responsabilité de G. Dostie contenant plus de 45 heures d'enregistrements ; transcriptions en accès libre, <http://pages.usherbrooke.ca/cfpq/corpus.php> (05/11/2013).
- Coseriu, E., 1981, « Los conceptos de < dialecto >, < nivel > y < estilo de lengua > y el sentido propio de la dialectología » [1958], *Lingüística española actual*, 3/1, p. 1-32.
- DHFQ : Poirier, C. (éd), 1998, *Dictionnaire historique du français québécois*, Sainte-Foy : PU Laval.
- DictBelg : Francard, M. et al., 2010, *Dictionnaire des belgicisms*, Bruxelles : De Boeck-Duculot.
- DLF : Valdman, A. et al., 2010, *Dictionary of Louisiana French*, Jackson : UP of Mississippi.
- DRF : Rézeau, P. (éd), 2001, *Dictionnaire des régionalismes de France (DRF)*, Bruxelles : De Boeck-Duculot.
- DRFA : Rézeau, P., 2007, *Dictionnaire des régionalismes du français en Alsace*, Strasbourg : Presses universitaires de Strasbourg.
- DSR : Thibault, A., 1997, *Dictionnaire suisse romand*, Genève : Zoé.
- ESLO, *Enquête Sociolinguistique à Orléans*. Corpus oral constitué de 700 heures d'enregistrements menés à Orléans (1968-1971 ; 2008-2013), en accès libre sur demande avec signature d'une convention ; accès direct en ligne à un sous-corpus anonymisé, <<http://eslo.tge-adonis.fr>> (19/11/3013).
- FEW : Wartburg, W. von, 1928-2003, *Französisches Etymologisches Wörterbuch*, 25 vol., Bonn et al. : Klopp et al.
- FLI, *Fichier lexical informatisé*. Base de données du français québécois du XVI^e au XXI^e siècles comportant 400 000 fiches, en accès libre, <<http://www.tlfq.ulaval.ca/fichier/>> (05/11/2013).
- Foucault, A. et J.-F. Raoult, 2010, *Dictionnaire de géologie*, 7^e édition [1980], Paris : Dunod.
- Frantext, Base de données textuelles de la littérature française : corpus à dominante littéraire constitué de quelque 248 millions de mots du XVI^e au XXI^e siècles, <<http://www.Frantext.fr/>> (01/05/2013).
- Genette, G., 1987, *Seuils*, Paris : Seuil.
- Grevisse, M., 2008, *Le Bon Usage : Grammaire française*, refondu par A. Goosse, 14^e édition, Paris / Louvain-la-Neuve : De Boeck-Duculot.
- Hansard. Corpus de français des Débats de la Chambre des communes au Canada, 1994-2013, consultable par mots-clés et séquences exactes et exclusion de mots-clés ; partiellement en accès libre, <<http://www.parl.gc.ca/HouseChamberBusiness/ChamberSittings.aspx>> (02/08/2013, 05/11/2013).
- I-FR, Corpus francophone de Leeds d'environ 200 millions de mots tirés de 50 000 pages du Web francophone en 2006 sans limitation de domaines, consultable avec recherches de concordances et de collocations, en accès libre, <Corpus.Leeds.ac.uk/internet.html> (01/05/2013).
- ILQ. *Index lexicologique québécois* : Inventaire des mots du français québécois ayant fait l'objet d'un commentaire ou d'une étude depuis 1750, établi sous la direction de Cl. Poirier et L. Mercier, 1979-1986 ; en accès libre <<http://www.tlfq.ulaval.ca/ilq/>> (05/11/2013).
- Maingueneau, D., 2009, *Les termes clés de l'analyse du discours* [1996], Paris : Seuil.
- , 2005, *Analyser les textes de communication* [1998], Paris : A. Colin.

Version d'auteur antérieure de : Wissner, Inka, 2015, « Du métadiscours au dictionnaire. Pour un étiquetage unifié des mises en relief en lexicographie francophone », in : Gaétane Dostie et Pascale Hadermann (edd.), *La dia-variation en français actuel. Études sur corpus, approches croisées et ouvrages de référence* (coll. Sciences pour la communication), Berne et al. : Peter Lang, 297-317, 21 pages.

- Poirier, P., 1993, *Le Glossaire acadien [1927-1933]*, Moncton : Editions d'Acadie.
- PR : Rey-Debove, J. et al. (éds), 2008, *Le Nouveau Petit Robert de la langue française 2009 [1967]*, Paris : Le Robert.
- Rey-Debove, J., 1997, *Le métalangage : étude linguistique du discours sur le langage*, 2^e édition [1978], Paris : Colin/Masson.
- RichessesRéz 2007 : Rézeau, P. (éd), 2007, *Richesses du français et géographie linguistique*, vol. I, Bruxelles : DeBoeck-Duculot.
- RichessesThib 2008 : Thibault, A. (éd), 2008, *Richesses du français et géographie linguistique*, vol. II, Bruxelles : De Boeck-Duculot.
- Rob : Rey, A., 2001, *Le Grand Robert de la langue française*, 2^e édition, Paris : Le Robert.
- Thibault, A., 2005, « Exemples linguistiques vs exemples métalinguistiques dans le *Dictionnaire suisse romand* et le *Dictionnaire des régionalismes de France* », in : M. Heinz (ed.), *L'exemple lexicographique dans les dictionnaires français contemporains*, Tübingen : Niemeyer, p. 95-113.
- TLF, 1971-1994, *Trésor de la langue française. Dictionnaire de la langue du 19^e et 20^e siècle (1789-1960)*, Paris : Gallimard.
- Usito : Cajole-Laganière, H. et al. (éds), 2013, *Usito. Parce que le français ne s'arrête jamais*, Sherbrooke : Delisme, accès payant en ligne, <www.usito.com> (02/08/2013).
- Wissner, I., 2012, « Les grands corpus du français moderne : des outils pour étudier le lexique diatopiquement marqué ? », *SKY Journal of Linguistics*, 25, p. 233-272, <<http://www.linguistics.fi/julkaisut/sky2012.shtml>> (02/08/2013).