

Lexicographie du français médiéval et corpus informatisé : le traitement lexicographique de verbes d'opinion et de connaissance à partir du corpus du DMF

Corinne Féron

**Édition électronique**

URL : <http://corpus.revues.org/1606>
ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 10 novembre 2008
ISSN : 1638-9808

Référence électronique

Corinne Féron, « Lexicographie du français médiéval et corpus informatisé : le traitement lexicographique de verbes d'opinion et de connaissance à partir du corpus du DMF », *Corpus* [En ligne], 7 | 2008, mis en ligne le 13 novembre 2009, consulté le 02 octobre 2016. URL : <http://corpus.revues.org/1606>

Ce document a été généré automatiquement le 2 octobre 2016.

© Tous droits réservés

*Lexicographie du français médiéval et corpus informatisé : le traitement lexicographique de verbes d'opinion et de connaissance à partir du corpus du DMF**

Corinne Féron

Introduction

- 1 L'utilisation de corpus informatisés a entraîné dans le domaine de la lexicographie des changements considérables. La lexicographie du moyen français est également concernée : c'est sur un corpus informatisé qu'est fondé en grande partie le *Dictionnaire du moyen français* (DMF), en cours de rédaction¹.
- 2 Dans un premier temps (DMF1), des lexiques « préalables » ont été rédigés, pour la plupart à partir de textes de ce corpus² ; la deuxième phase (DMF2) a permis d'ajouter de nouveaux lexiques, dont un lexique complémentaire qui traite de mots absents des autres lexiques³ et étend ainsi notablement la nomenclature⁴. L'étape actuelle (DMF3) consiste à rédiger des articles synthétiques⁵ pour des mots fréquents ; le rédacteur se fonde sur le travail déjà réalisé – les articles des lexiques – et sur le corpus informatisé⁶ ; il dispose en outre d'une riche documentation papier, comprenant notamment les glossaires d'un grand nombre d'éditions critiques.

Le fait de disposer d'un corpus vaste et varié permet d'enrichir considérablement les articles de dictionnaire : Trotter (2000) l'a montré en comparant un article de la première édition de l'*Anglo-Norman Dictionary* à l'article correspondant de la seconde édition. Pour le DMF, nous ne pouvons pas procéder à ce type de comparaison : il n'existe pas de dictionnaire portant précisément sur le moyen français auquel nous pourrions comparer

les articles déjà rédigés. Nous examinerons donc simplement le rôle que joue le corpus des textes intégraux dans la préparation, puis dans la rédaction d'un article :

- d'une part, comment, au cours de la rédaction d'un article, le rédacteur peut-il tirer profit de ce corpus ? A quelles étapes de son travail ?
 - d'autre part, en quoi le recours à un corpus informatisé permet-il de répondre aux exigences fixées pour le DMF (Martin 1997 ; Martin s.d.) ? Ces exigences concernent à la fois l'exhaustivité (qui passe par le recensement, pour un mot donné, de toutes ses acceptions et de ses constructions⁷) et l'adéquation de la description : sont concernées ici les définitions, les conditions d'emplois – par exemple, pour les verbes, la sous-catégorisation sélectionnelle, voire l'indication de l'hyperonyme des noms en fonction de sujet et d'objet –, ainsi que l'organisation de l'article, qui vise, idéalement, à donner une image globale du fonctionnement des polysèmes.
- 3 Il sera question ici du traitement lexicographique de verbes qui, dans certaines de leurs acceptions au moins, relèvent de l'opinion ou de la connaissance (nos exemples concerneront *croire*, *cuidier*⁸, *penser*⁹, *savoir* et *connaître*). Pour ce travail, nous avons consulté le corpus du DMF par le biais de Frantext¹⁰, qui donne accès à 218 textes¹¹ (soit 6 790 794 occurrences), dont le plus ancien date de 1330, le plus récent de 1502.
 - 4 Il semble aller de soi que le recours à un corpus informatisé permet de répondre à l'exigence d'exhaustivité fixée pour le DMF : recenser toutes les acceptions, toutes les constructions effectivement présentes dans le corpus. La taille de ce corpus peut en outre laisser espérer qu'une part importante des types d'emplois du moyen français¹² y est attestée, notamment ceux qui relèvent du lexique courant ; les mots ou emplois propres aux lexiques spécialisés y sont moins bien représentés¹³, les textes littéraires étant majoritaires dans le corpus¹⁴.
 - 5 Pour un verbe donné, on peut rassembler toutes les occurrences à partir de la liste des formes de ce verbe ; celle-ci peut être importée sur Frantext à partir du DMF¹⁵. Après y avoir supprimé, si nécessaire, les formes n'appartenant pas au paradigme verbal, le rédacteur constitue son corpus d'étude, par une requête simple portant sur l'ensemble des formes de la liste.
 - 6 L'examen de l'ensemble des occurrences¹⁶ permet de repérer tous les emplois, mais c'est à l'évidence une tâche longue et fastidieuse lorsque les corpus comprennent plusieurs milliers d'occurrences (3362 pour *croire*, 4524 pour *penser*, 2297 pour *cuidier*, 4116 pour *connaître*, 15778 pour *savoir*). Or, Frantext comprend des fonctionnalités (étude de voisinage, recherche de cooccurrences) qui aident à exploiter ces corpus ; nous verrons quel parti peut être tiré de ces services, en faisant porter notre attention sur deux aspects importants dans la description lexicographique des verbes : l'environnement lexical – incluant les actants ou circonstants aussi bien que des cooccurrents non dépendants du verbe – (section 1), puis le recensement des constructions du verbe (section 2).

1. L'environnement lexical

1.1. Les « portraits de mots »

- 7 L'atout des corpus informatisés le plus souvent mis en avant est la possibilité de repérer la combinatoire des mots : les rédacteurs du TLF disposaient ainsi pour les mots lexicaux d'un recueil des groupes binaires, « sortes d'inventaires des distributions et des

collocations stabilisées ou fréquentes » (Quemada 1983 : 113). Il existe désormais des outils permettant d'obtenir la liste des sujets, objets et modificateurs typiques (Béjoint 2007 : 12-14) ; on établit ainsi des « portraits de mots », fournissant, dans le cas des verbes, des informations sur l'environnement lexical et sur les constructions.

- 8 Il s'agit là d'une base précieuse pour le rédacteur, tant pour l'organisation de son article que pour le travail définitoire ; en particulier, la récurrence de tel objet, tel sujet ou circonstant peut être révélatrice des propriétés du mot (ou d'une acception de ce mot), ou encore permettre de repérer des collocations ou des locutions.
- 9 Les informations recueillies de cette façon ont leur place dans l'article de dictionnaire : elles renseignent sur les usages fréquents du mot-entrée – ce qui paraît indispensable tant dans un dictionnaire d'encodage que dans un dictionnaire de décodage ; de plus, elles font du dictionnaire un outil utilisable pour des procédures d'étiquetage de corpus, notamment dans le cas d'unités polysémiques (Véronis 2004, Martin 2001).

1.2. L'étude de voisinage dans Frantext

- 10 En ce qui concerne Frantext, l'étude de voisinage permet d'obtenir, par ordre alphabétique ou par ordre croissant ou décroissant des fréquences, les cooccurrents d'une forme donnée (ou d'une liste de formes). Cette fonctionnalité, qui fournit des listes de formes – et non des concordances –, est particulièrement utile pour l'étude d'un état de langue du passé, pour lequel nos intuitions sur les cooccurrences fréquentes, les collocations ou les locutions sont défailantes. Néanmoins, l'interprétation des résultats est assez malaisée, pour trois raisons.
- 11 *La difficulté principale tient à l'absence d'outil de lemmatisation : d'une part, le pivot inclut des formes qui peuvent ne pas appartenir au paradigme verbal (cf. note 17) ; d'autre part, la haute fréquence d'un mot cooccurrent peut être masquée si celui-ci apparaît dans le voisinage du pivot sous de multiples formes (ce qui est notamment le cas des verbes cooccurrents).
- 12 *Pour les mots grammaticaux ou ceux qui sont susceptibles de fonctionner comme mots grammaticaux, les graphies modernes n'apparaissent pas dans les résultats de la requête¹⁷ : ainsi, les formes *seur*, *del*, *quant*, *quë* figurent dans les listes, mais non *sur*, *du*, *quand*, *que* ; il en est de même pour certains verbes aptes à fonctionner comme auxiliaires ou semi-auxiliaires (*peult* et *faict* apparaissent, mais ni *peut* ni *fait*) voire pour certains adverbes (la graphie *mieux* se trouve dans les listes, mais non *mieux*).
- 13 *Les formes figurant dans la dépendance du pivot sont mêlées à des formes qui ne le sont pas, voire qui appartiennent à une autre phrase¹⁸.
- 14 Nous donnons dans les tableaux 1 et 2 un aperçu des résultats obtenus pour les cinq verbes étudiés. On a limité le voisinage à 5 mots avant et 5 mots après le pivot¹⁹ ; la liste obtenue (par fréquence décroissante) a ensuite été triée manuellement selon la classe grammaticale des formes cooccurrentes (compte non tenu des mots grammaticaux – dont seules les graphies anciennes figurent dans les listes – et des adjectifs, puisqu'ils ne sont pas directement en relation avec le verbe). Les formes pouvant relever de plusieurs classes grammaticales, par exemple *mal* ou *mort*, ont été placées (après vérification dans le corpus) dans la classe à laquelle elles appartiennent le plus souvent lorsqu'elles sont dans le voisinage du pivot.

Tableau 1 : formes nominales

	<i>croire</i>	<i>cuidier</i>	<i>penser</i>	<i>connaître</i>	<i>savoir</i>
1	267 dieu	80 roy	216 cuer	167 dieu	260 dieu
2	238 rente	60 gens	154 dieu	134 choses	222 dame
3	234 cens	56 dieu	144 dame	82 serement	202 sire
4	194 conseil	48 femme	113 desir	77 verité	189 roy
5	70 foy	42 verité	104 mal	61 gens	113 homme

Tableau 2 : formes verbales (sauf auxiliaires et semi-auxiliaires²⁰)

	<i>croire</i>	<i>cuidier</i>	<i>penser</i>	<i>connaître</i>	<i>savoir</i>
1	28 tient	76 trouver	57 mettre	188 confessa	101 parler
2	24 amer	72 prendre	50 aler	42 scet	101 parler
3	21 penser	64 mort	47 parler	34 veu	80 aler
4	16 retenu	55 ferir	43 garder	28 afferma	59 penser
5	15 entendre 15 prins 15 servir	50 venir	35 ymaginer	27 veoir	57 trouver

- ¹⁵ Pour *croire*, la présence de *rente* et *cens* (« redevance ») dans le tableau des formes nominales s'explique par l'absence de lemmatisation : la liste des formes du pivot inclut *crois*, qui peut être une forme verbale ou une forme nominale ; c'est alors soit une graphie de *croix*, soit le déverbal de *croître* et dans ce dernier cas, *crois* est en cooccurrence avec *cens*²¹ et *rente* (toutes ces occurrences se trouvant dans un texte à caractère juridique, les *Chartes de l'abbaye de Saint-Magloire*).
- ¹⁶ Actuellement, les listes fournies par l'étude de voisinage ne fournissent donc qu'une image très floue de la combinatoire lexicale des mots ; cependant, elles attirent l'attention du rédacteur sur des cooccurrences qu'il aurait pu ne pas repérer lors d'un examen « manuel » de l'ensemble des occurrences : il peut alors procéder à une recherche dans le corpus (par le biais de la recherche de cooccurrences) pour vérifier la pertinence de ses premières observations. Nous donnerons des exemples concrets de cette démarche à propos de quelques-unes des formes nominales et verbales qui apparaissent en tête des listes de voisinage. Dans chaque cas, les différentes graphies appartenant à l'unité retenue ont été regroupées (pour *ymaginer* par exemple, on a tenu compte des 11 autres

formes qui se trouvent également dans la liste des cooccurrents de *penser – imaginer, ymaginera*, etc. –, ce qui porte le total d'occurrences pour ce verbe à 51).

1.3. Formes nominales fréquentes

- 17 Si l'on observe les contextes où le nom dieu (sous cette forme ou sous une forme de cas sujet : *dieux, dieus, diex*) est dans le voisinage des cinq verbes, on constate que cette cooccurrence ne présente guère d'intérêt pour l'étude de *cuidier – dieu* n'est jamais dans la valence de ce verbe –, ni pour *penser* et *connaître* : *dieu* est certes parfois objet²² ou sujet de ces verbes, mais ces cooccurrences ne permettent pas de délimiter des acceptions ou des constructions particulières. La présence de ce nom est en revanche significative pour *croire*, en ce sens qu'elle correspond le plus souvent à une acception spécifique du verbe²³ : « avoir foi en l'existence de (Dieu) et mettre sa confiance en lui »²⁴.
- 18 Il en est de même pour *savoir*, mais pour une autre raison. Si l'on recherche les contextes correspondants en procédant en deux temps, selon la place du nom par rapport au verbe, on constate que *dieu* est le plus souvent avant le verbe (les cas où *dieu* précède *savoir* sont deux fois et demi plus nombreux que ceux où *savoir* précède *dieu*) ; parmi les contextes concernés, quelques schémas récurrents se dégagent :
- (i) *Dieu sait* + subordonnée conjonctive en *que* ou + interrogative (totale ou partielle) ;
 - (ii) *Dieu sait* + pronom anaphorique (*Dieu le sait, ce sait Dieu*) ;
 - (iii) *Dieu sait* (en incise) ;
 - (iv) *Dieu sait quand/comment* fonctionnant comme adverbe ;
 - (v) *Dieu sait quel...* fonctionnant comme déterminant.
- 19 En (iv) et (v), nous avons clairement affaire à des locutions : la séquence nom + verbe + mot interrogatif fonctionne dans la phrase comme un mot simple – comme un adverbe ou comme un déterminant. En (i) à (iii), en revanche, on a une structure phrastique ordinaire en sujet + verbe + complément (effacé dans le cas de l'incise) ; *savoir* y est employé dans une acception courante (« avoir présent à l'esprit (un objet de pensée qu'on identifie et qu'on tient pour vrai) ») et dans des constructions bien attestées avec d'autres sujets. Dans les cas (ii) et (iii), la séquence fonctionne comme modalisateur : le locuteur donne du poids à son assertion en y insérant un commentaire qui renvoie à la connaissance que Dieu a des faits. Dans le type (i), par exemple dans l'extrait suivant (avec conjonctive en *que*) :
- Quant le povre deduit du lit Est passé par aucunes nuis, Lors te saudront les grans ennuis, Car tu ne pourras achever Son delit sanz ton corps grever, Qui adonc reposer vouldras ; Mais *Dieux scet que* tu ne pourras Rendre le deu qu'elle demande Quant au delit. (E. Deschamps, *Le miroir de mariage*, c.1385-1403, p. 54)
- 20 la stratégie argumentative est un peu différente : en énonçant *Dieu sait que P*, le locuteur ne prend pas en charge lui-même *P*, ne l'asserte pas à son propre compte ; il s'en remet à Dieu et, ce faisant, donne *P* pour incontestable ; par ailleurs, il y a ici un décalage entre la structure syntaxique et le fonctionnement sémantique et discursif de la séquence (cf. Apothéloz 2003 : 250 à propos de *je crois que*) : si *Dieu sait* est en apparence la régissante, c'est la complétive qui constitue le contenu asserté (et commenté par *Dieu sait*).
- 21 Au total, la haute fréquence d'un nom dans le voisinage d'un verbe (du moins dans les listes que nous avons obtenues) peut être d'intérêt très variable : en ce qui concerne le nom *dieu*, elle est négligeable pour trois des verbes étudiés, elle correspond à une

acceptation pour l'un des verbes (*croire*), et enfin, pour *savoir*, à des locutions où à des emplois quasi-locutionnels.

- 22 Pour le substantif *mal*, fréquent dans le voisinage de *penser*, on observe que, le plus souvent, il est non déterminé ou bien précédé d'un déterminant indéfini (*nul, aucun*) ; on a alors soit une construction directe : (y) *penser mal*²⁵, soit une construction indirecte : (y) *penser à mal*. Dans chaque cas, deux interprétations sont possibles selon les contextes²⁶ : (y) *penser (à) mal* peut signifier « former en soi (une intention mauvaise) » ou « juger, estimer (qu'il y a qqc. de mal) ».
- 23 Or, dans le voisinage de *penser* se trouvent aussi un certain nombre de noms qui désignent un comportement moralement blâmable – qui sont donc sémantiquement assez proches de *mal* –, avec une fréquence relativement élevée pour certains (*laidure* : 25, *folie* : 21, *deshonneur* : 11, *malice* : 11, *honte* : 8, *tricherie* : 7, *outrage* : 6...); comme *mal*, ils sont construits directement ou indirectement, mais seul le sens « former en soi (une intention) » est alors attesté. La liste comporte aussi des antonymes : *honneur, loyauté*, etc. ; beaucoup plus rares, ces noms n'apparaissent que dans la construction directe et *penser y* signifie toujours l'intention.
- 24 On voit que ce qu'il est possible de faire sur un petit corpus²⁷ par un examen exhaustif des occurrences peut être fait sur un corpus plus vaste grâce à la liste de voisinage : repérer des acceptations et des constructions du verbe et en outre faire la liste des noms que le verbe peut avoir comme objet dans chacun des cas.

1.4. Formes verbales fréquentes

- 25 Parmi les formes verbales les plus fréquentes dans le voisinage des verbes étudiés, certaines n'appellent guère de commentaires : c'est le cas de celles que l'on trouve fréquemment à proximité de *cuidier* : *trouver, prendre, ferir* et *venir* ; aucun lien sémantique n'est perceptible entre ces verbes²⁸ et *cuidier* et, de fait, on constate – lorsqu'on cherche dans le corpus les contextes correspondants – qu'ils sont soit sans lien syntaxique avec *cuidier*, soit régis par *cuidier* (au sens de « penser », de « avoir l'intention de » ou de « être sur le point de ») ; aucune de ces cooccurrences n'est significative du sens de ce verbe.
- 26 Pour *croire* en revanche, figurent dans la liste des cooccurrents de rang 1 à 5 des verbes qui lui sont sémantiquement apparentés : *servir* et *amer* s'associent à *croire* au sens de « s'en remettre à qqn » (l'objet étant soit *Dieu*, soit, dans la poésie courtoise, *dame*) ; quant à *tenir*²⁹ et *penser*, ils s'apparentent à *croire* au sens de « tenir pour vrai » ou « avoir la conviction, la certitude (que...) » et lui sont souvent coordonnés³⁰.
- 27 Lorsqu'on travaille, comme c'est le cas ici, sur des unités lexicales qui, pour certains au moins de leurs emplois, appartiennent à une même classe sémantique (*savoir* et *connaître* d'une part, *penser, croire* et *cuidier* d'autre part), l'examen des cooccurrences permet en outre de préciser les différences et les similitudes entre ces unités : ainsi, pour les verbes susceptibles de relever de l'opinion, on note que la forme verbale qui est le plus souvent à proximité de *croire* est *tenir* ; or, *tenir* peut aussi être coordonné à *cuidier*, mais jamais à *penser*. Cette divergence dans la combinatoire des verbes serait un indice à retenir pour un réexamen de la « synonymie » entre les verbes *croire, cuidier* et *penser*.
- 28 Au total, si elle ne permet pas de faire des « portraits » de mots, l'étude de voisinage offre un réel intérêt pour le rédacteur (pourvu, il est vrai, qu'il prenne le temps de regrouper

toutes les formes des mots cooccurrents qui semblent significatifs et qu'il examine les contextes, par le biais de recherche de cooccurrences).

- 29 – On a vu qu'il est possible d'une part de repérer des sujets ou objets fréquents, dont certains relèvent du figement (*Dieu sait...*) ; mais l'intérêt majeur de l'étude de voisinage est sans doute de pouvoir établir le paradigme des noms figurant en fonction d'objet pour une acception donnée ; à défaut de faire apparaître dans l'article la liste des noms recensés, on peut y donner des indications précises sur la classe à laquelle ils appartiennent et sur leur détermination (déterminant zéro ou déterminant indéfini dans l'exemple que nous avons pris), et enfin préciser éventuellement dans quels textes ces emplois se trouvent majoritairement (par exemple, la plupart des occurrences de *penser* (à) *mal / laidure*, etc. proviennent de l'œuvre de Machaut).
- 30 – D'autre part, les verbes cooccurrents conduisent à dégager au moins certaines des acceptions des verbes étudiés et, dans le cas où l'on étudie parallèlement des verbes ayant quelque parenté sémantique, de noter des différences dans leur combinatoire lexicale.

2. Les constructions verbales

- 31 Le repérage des constructions est une phase importante dans le travail du rédacteur : d'une part, parce qu'il est chargé de rendre compte dans l'article des usages du mot traité, donc de son intégration syntagmatique ; d'autre part, parce que lorsqu'il y a³¹ coïncidence entre sens et construction, c'est un moyen pour structurer clairement l'article. À défaut de pouvoir réaliser cette tâche de façon automatique, le rédacteur peut obtenir, à l'aide de la recherche de cooccurrences, des listes d'occurrences relevant d'une construction donnée³² pour en observer si nécessaire les propriétés ou simplement en extraire des exemples pour son article.
- 32 Ce travail ne peut toutefois se faire que pour des constructions mettant en jeu des morphèmes grammaticaux sous forme de chaînes de caractères ; en l'absence d'étiquetage morphologique du corpus de moyen français, il est impossible d'obtenir par Frantext toutes les occurrences d'un verbe ayant dans son environnement par exemple un infinitif ou un substantif.
- 33 Ces requêtes peuvent se faire à partir d'une liste de constructions établie à partir des dictionnaires existants : nous n'avons pas de dictionnaire consacré exclusivement au lexique des XIV-XV^e siècles, mais nous disposons des « lexiques préalables » d'auteurs ou de genres (voir *supra*) ; il semble par ailleurs indispensable de tenir compte des constructions attestées dans les textes de la période précédente et de la période qui suit : le TL pour l'ancien français³³, le Huguet pour le français du XVI^e siècle, et enfin le Godefroy, qui englobe ancien et moyen français.
- 34 Pour le verbe *penser* par exemple, on obtient ainsi la liste suivante (on ne retient que les constructions marquées par un mot grammatical) :
- penser que
 - *penser* + mot interrogatif
 - *penser se* (conjonction introduisant une interrogative indirecte)
 - penser à / en / de / sur
 - soi penser
 - soi penser que
 - *soi penser* + mot interrogatif

- soi penser se
 - soi penser de
- 35 La difficulté essentielle pour la formulation des requêtes dans le corpus concerne la distance « idéale » entre le verbe et le morphème grammatical. Ainsi, pour obtenir les occurrences de *penser* régissant une complétive en *que*, si l'on fixe une distance maximale de 3 entre la forme verbale et la conjonction³⁴, on obtient des occurrences non voulues, telles que *penser à celle que...*³⁵ ; cependant, en réduisant cette distance on exclut nombre d'occurrences pertinentes, notamment celles où s'intercale, entre le verbe et la conjonction, un syntagme nominal sujet (« Or pourroient par adventure *penser* aucunes personnes *que* Sarre eust par mal et par envye enchassé Agar sa meschine et Ysmael son filz », *Le ménagier de Paris*, 1394, p. 62) ou encore une locution adverbiale (*penser en vérité / de vrai que...*).
- 36 Souvent, les résultats sont longs à exploiter : par exemple, pour la construction *penser de* (la requête portant sur l'ensemble du paradigme de *penser* et sur toutes les formes que peut prendre la préposition : *de, d', del, dou, du, des*)³⁶, on obtient 1139 occurrences, même en ne prévoyant que deux mots entre la forme verbale et la préposition. Il paraît en revanche réaliste d'examiner dans son entier la liste des cas où *penser* régit une subordonnée interrogative totale : la requête portant sur la cooccurrence de *penser* et de *se/s'* donne 119 occurrences (parmi lesquelles certaines sont non pertinentes).
- 37 En ce qui concerne la construction *soi penser se*, qui n'apparaît que dans un exemple du TL (t. VII : 677), on peut rapidement vérifier qu'elle est également très rare en moyen français : en formulant une requête portant sur la cooccurrence de *penser*, de la conjonction *se* et d'un pronom réfléchi, on obtient 24 occurrences, qu'il suffit de parcourir pour constater qu'une seule correspond à la construction recherchée.
- 38 L'objectif du rédacteur n'est pas de quantifier, mais il est néanmoins important qu'il puisse se faire une idée de la vitalité des différentes constructions, afin de pouvoir signaler celles qui sont rares, qu'il s'agisse de constructions archaïques, nouvelles, ou simplement marginales.

Conclusion

- 39 Notre attention s'est moins portée sur l'atout que représente en lui-même le corpus informatisé pour le DMF que sur son exploration en vue de la rédaction d'articles. On a vu que les fonctionnalités qu'offre actuellement Frantext facilitent le travail du rédacteur : ils lui permettent notamment de repérer des cooccurrences lexicales fréquentes (par l'étude de voisinage) ou d'interroger le corpus sur certaines constructions. C'est là un apport important pour un dictionnaire qui vise à donner des descriptions explicites des usages – ce qui n'est pas toujours le cas dans les dictionnaires de français médiéval antérieurs³⁷ (cf. à propos des locutions le relevé de Trotter (2004 : 586) concernant le nom *grâce*).
- 40 Toutefois, bien souvent, le travail nécessite la consultation de longues listes et un tri manuel. De plus, à moins de se résigner à consulter l'ensemble des occurrences, certains emplois risquent d'échapper : c'est le cas par exemple pour *croire* + infinitif, dont le corpus du DMF offre pourtant – à notre connaissance – les premières occurrences (dans un dit de Guillaume de Machaut). La lemmatisation du corpus simplifierait l'utilisation de l'étude de voisinage et l'étiquetage morphologique du corpus autoriserait des requêtes

aujourd'hui impossibles (*croire* + infinitif par exemple); au-delà, un étiquetage sémantique permettrait par exemple de retrouver dans le corpus toutes les occurrences de *penser* de suivi d'un nom désignant un animal (et donc de réunir bon nombre des occurrences de *penser* au sens de « s'occuper de ») ou de *savoir* + nom /concret/ (*savoir le chemin, la voie, le sentier, le pays, la terre, etc.*). Des textes de français médiéval sont d'ores et déjà étiquetés (Marchello-Nizia 2004 : 60-61) mais les difficultés sont nombreuses (Prévost et Heiden 2002 ; Prévost 2005 : 169).

- 41 On rappellera enfin que, si les outils d'exploitation du corpus ne sont pas encore aussi efficaces qu'on pourrait le souhaiter, le rédacteur dispose d'un autre type d'« outil », à savoir les dictionnaires existants : il est facile d'en remarquer les manques – comme nous l'avons fait ci-dessus –, il n'en reste pas moins qu'ils peuvent être le point de départ de requêtes fructueuses, concernant certaines constructions verbales (on en a donné un exemple ci-dessus à propos de *penser*), des syntagmes (par ex. *vos ne savez*, dans le TL 254 : une recherche dans le corpus du DMF fournit rapidement quelques exemples de ce type) ou encore des cooccurrents fréquents (*penser* + adverbe ou syntagme à valeur temporelle : *tote jor, longues, longuement, et de nuit et de jor ...* dans le TL).

BIBLIOGRAPHIE

Dictionnaires

AND : *Anglo-Norman Dictionary*, sur le site Anglo-Norman On-line Hub : <http://www.anglo-norman.net>.

DMF : *Dictionnaire du Moyen Français*, ATILF / Nancy Université - CNRS. Site internet : <http://www.atilf.fr/dmf>.

FEW : Wartburg W. von (1930-). *Französisches etymologisches Wörterbuch*. J.C.B. Mohr.

GD : Godefroy F. (1880-1902), *Dictionnaire de l'ancienne langue française du IX^e au XV^e siècle*. Paris : Wieweg et Bouillon.

Huguet : Huguet E. (1925-1967). *Dictionnaire de la langue française du seizième siècle*. Paris : Champion [puis] Didier.

TL : Tobler A. et Lommatzsch E. (1925-2002). *Altfranzösisches Wörterbuch...* Wiesbaden/Stuttgart : Franz Steiner.

TLF : *Trésor de la langue française*. Paris : éd. du CNRS [puis] Gallimard, 1971-1994.

Références bibliographiques

Apothéloz D. (2003). « La rection dite "faible" : grammaticalisation ou différentiel de grammaticité ? », *Verbum* XXV-3 : 241-262.

Béjoint H. (2007). « Informatique et lexicographie de corpus : les nouveaux dictionnaires », *Revue française de linguistique appliquée* XII-1 : 7-23.

Lavis G. (1977). « Remarques sur les valeurs syntactico-sémantiques des verbes a.fr. "penser", "cuidier", "croire" suivis d'un infinitif ou d'un substantif objet direct », in A. Varvaro (éd.), *Atti del*

- XIV congresso internazionale di linguistica e filologia romanza, IV. Naples/Amsterdam : G. Macchiaroli / J. Benjamins, 467-487.
- Marchello-Nizia C. (2004). « Linguistique historique, linguistique outillée : les fruits d'une tradition », *Le Français moderne* 72-1 : 58-70.
- Martin R. (1992). « L'article tenir du Dictionnaire du moyen français (DMF) », *Vox Romanica* 51 : 129-163.
- Martin R. (1997). « Les « normes » du DMF (Dictionnaire du moyen français) », in B. Combettes et S. Monsonégo (éds.), *Le moyen français. Philologie et linguistique : approches du texte et du discours*, Actes du VIII^e colloque international sur le moyen français. Paris : Didier érudition, 297-305.
- Martin R. dir. (1998). *Dictionnaire du moyen français*, vol. A-AH, prépublication non commercialisée, s.l. (CNRS, INALF).
- Martin R. (2001). *Sémantique et automate*. Paris : PUF.
- Martin R. (s.d.) Présentation du DMF. <http://atilf.atilf.fr/gsouvay/dmf2/PresentationDMF2.pdf>
- Prévost S. (2005). « Constitution et exploitation d'un corpus de français médiéval : enjeux, spécificités et apports », in A. Condamines (dir.), *Sémantique et corpus*. Paris : Lavoisier, 147-176.
- Prévost S. et S. Heiden (2002). « Etiquetage d'un corpus de français médiéval : enjeux et modalités », in C. D. Pusch et W. Raible (éds.), *Romance Corpus Linguistics – Corpora and Spoken Language*. Tübingen : Gunter Narr Verlag, 126-137.
- Quemada B. (1983). « Bases de données informatisées et dictionnaires », *Lexique 2* : 101-120.
- Trotter D. (2000). « L'avenir de la lexicographie anglo-normande : vers une refonte de l'Anglo-Norman Dictionary ? », *Revue de linguistique romane* 64 : 391-407.
- Trotter D. (2004). « Le problème de l'identification des locutions dans une langue morte : l'exemple de l'Anglo-Norman Dictionary », in M. Colombo Timelli et C. Galderisi (éds.), « Pour acquérir honneur et pris ». *Mélanges de Moyen Français offerts à Giuseppe Di Stefano*. Montreal : CERES, 583-592.
- Véronis J. (2004). « Quels dictionnaires pour l'étiquetage sémantique ? », *Le français moderne* 72-1 : 27-38.

NOTES

- *. Nous suivons les « rectifications » orthographiques de 1990 – d'où par exemple l'absence d'accent circonflexe sur *connaitre*, *apparaître*, etc.
1. Sur les différentes phases de la rédaction du DMF, voir Martin (s.d.).
 2. Ils ont ensuite été lemmatisés ; on peut ainsi obtenir, pour un mot donné, l'ensemble des articles correspondants.
 3. Ces lexiques excluent les mots grammaticaux ; toutefois, le choix des mots traités étant du ressort des rédacteurs, on observe des divergences pour les mots à la limite du grammatical et du lexical, par exemple pour certains adverbes (ainsi *moult* est présent dans cinq des 16 lexiques d'auteurs ou de « genres ») et pour les prépositions lexicales (*selon* a été retenu dans un lexique, *jouxte* dans deux).
 4. Ces deux versions sont consultables sur internet : <http://www.atilf.fr/dmf>.
 5. La tranche a-ah (Martin dir. 1998) a été entièrement rédigée mais elle n'a pas été commercialisée ; par ailleurs, quelques articles ont donné lieu à des publications (par exemple, Martin 1992).

6. Les rédacteurs de l'ATILF disposent en outre d'un corpus informatisé constitué d'extraits de textes qui ont été jugés particulièrement intéressants.
7. Les indicateurs grammaticaux retenus renvoient aux « étiquettes » traditionnelles (par exemple pour les verbes : emploi transitif, intransitif, impersonnel, etc.), mais le rédacteur peut apporter des précisions sur les actants du verbe et leur construction par le biais de formules syntaxiques (*qqc. acointe qqn de qqc, sv acointier*, dans le Lexique de Machaut) et par des commentaires insérés dans des « entre-crochets ».
8. *Cuider* (*cuidier* en ancien français) signifie pour l'essentiel « tenir pour vrai ou probable », « avoir l'intention de » ou encore « être sur le point de » ; dans le premier cas, il régit une complétive en *que*, une « proposition infinitive » ou un infinitif ; dans les deux autres cas, on a *cuidier* + infinitif.
9. Ce verbe a aussi le sens de « prendre soin de » : les graphies du moyen français ne distinguent pas les sens référant à une activité intellectuelle de ceux qui renvoient à une activité matérielle.
10. http://www.atilf.fr/dmf/index_textes.htm.
11. Ce corpus est également interrogeable à partir du site DMF2, qui donne accès à quelques textes supplémentaires.
12. *I.e.* des emplois qui sont attestés dans l'ensemble des textes de cette période.
13. Voir Trotter (2000 : 400-402) pour un exemple concret de l'intérêt que présentent les textes non littéraires : dans la seconde édition de l'AND, fondée sur un corpus comprenant de nombreux textes techniques (juridiques et médicaux par exemple), l'article *abatre* donne des sens et des constructions que d'autres dictionnaires ne signalent pas ou dont ils ne fournissent que peu d'attestations.
14. Dans la pratique effective des rédacteurs cependant, les « lacunes » de ce corpus sont compensées par les occurrences fournies par la documentation papier et par la base des textes dépouillés sélectivement (voir *supra*, note 6).
15. Sur Frantext, la fonction « création d'une liste par flexion d'un verbe » fournit des listes, mais elles sont généralement incomplètes.
16. Parmi lesquelles se trouvent un certain nombre d'homonymes, par exemple *pensé* et *croyant* qui sont soit des participes, soit des substantifs.
17. L'étude de voisinage, conçue pour des textes de français moderne, ne concerne que les mots « lexicaux ».
18. Nous n'avons pas exploité la possibilité de faire porter l'étude de voisinage sur la phrase où se trouve le pivot : ce type de recherche ne permet pas de limiter l'observation à l'environnement immédiat du pivot ; pour cette raison, elle nous semble assez peu utile dans le cas des verbes, pour lesquels ce sont surtout les actants et circonstants récurrents que l'on a besoin de repérer.
19. Pour *savoir*, le programme a été interrompu en raison de l'importance du corpus ; les résultats fournis ne portent que sur les 10001 premiers voisinages.
20. Nous avons exclu les formes de *pouvoir*, *devoir*, *falloir*, *vouloir*, *faire*.
21. *Crois de cens* « rente nouvellement constituée », par opposition à la *rente primitive*.
22. Par exemple, en fonction d'objet : *penser Dieu*, *penser de/à Dieu* ; *connaître Dieu* ; en fonction de sujet : *Dieu pense de qqn ou qqc* (« Dieu prend soin, s'occupe de »), *Dieu connaît tout*.
23. Les textes littéraires semblent employer indifféremment la construction directe (*croire Dieu*) ou indirecte (*croire à / en Dieu*) – ou du moins, on ne parvient pas à saisir de différence sémantique entre ces constructions. En revanche, dans un traité de théologie (*La Somme abrégée de théologie*, 1481, p. 99) *croire Dieu* (« croire que Dieu ») est explicitement opposé à *croire à Dieu* (« croire à ses paroles, doctrines et enseignements ») et à *croire en Dieu* (« en créant Dieu le amer et aler à lui par amour »).
24. Dans la poésie lyrique, *croire* dans cette acception peut aussi avoir pour objet le nom *dame*.
25. Souvent dans un syntagme prépositionnel en *sans* : *sans (y) penser mal*.
26. Les deux interprétations sont parfois acceptables pour un même exemple.

27. Cf. par exemple Lavis (1977), qui, ayant examiné toutes les occurrences de *penser*, *croire* et *cuidier* dans les romans de Chrétien de Troyes, liste les objets nominaux que peut prendre chacun des trois verbes.
28. *Trouver que* n'est pas un verbe d'opinion en moyen français ; en revanche, on relève parfois *trouver qqn/qqc.* + attribut de l'objet au sens moderne.
29. *Tenir que* signifie souvent « considérer comme vrai » en moyen français.
30. De même, *confesser* et *affirmer* s'associent à *connaître* verbe déclaratif (« dire qu'on admet pour vrai »), *voir* (*veu*, *veoir*) et *savoir* (*scet*) à *connaître* verbe de connaissance ; pour *penser*, seules sont significatives les cooccurrences avec *garder* et *imaginer*, qui peuvent être coordonnés à *penser* respectivement au sens de « prendre soin de » et de « former dans son esprit l'image de (un objet, un projet) ».
31. Ce n'est pas toujours le cas : on a vu par exemple que *penser mal* / *penser à mal* sont parfois – autant qu'on puisse en juger – sémantiquement équivalents.
32. La recherche de cooccurrences est également possible par le biais du DMF2 (la requête pouvant être faite à partir des lemmes).
33. Ce dictionnaire cite aussi quelques exemples du XIV^e siècle.
34. Ce qui correspond aux cas où 2 mots au plus figurent entre le verbe et la conjonction.
35. On peut en revanche exclure les séquences dans lesquelles une virgule se trouve entre le verbe et la conjonction, et ainsi écarter certains au moins des cas où la conjonction n'est pas dans la rection du verbe – par exemple : « Mais Dieux, qui de ses amis *pense*, Fist *qu'il* fu en convalescence » (Machaut, *La prise d'Alexandrie*, 1369, p. 130).
36. La requête porte sur les cas où la préposition est à droite ou à gauche de *penser*, le syntagme prépositionnel pouvant être antéposé au verbe (« Mais *de* marchander *penser*as », *Miracle de un marchand et un larron*, 1349, p. 96).
37. Par exemple, le TL (t. VII : 669-670) cite une occurrence de *penser à mal*, suivi d'une traduction entre parenthèses et plusieurs de *penser felonie/folie/traison*, mais ne fait pas apparaître ces syntagmes clairement dans l'article ; quant au GD (t. VII : 87a et t. X : 313), il ne mentionne pas ce type d'emploi. En ce qui concerne les séquences en *Dieu sait*, on trouve *Dieu sait comment* (employé comme locution adverbiale) dans le GD (t. X : 636a) et *Dieu le sait* – suivi d'un exemple présentant la construction avec interrogative indirecte – dans le TL (t. IX : 253), qui mentionne par ailleurs *Dieu le sache*. Seul le FEW (t. XI : 195a) donne une liste complète des séquences en question (mais avec des dates qui seraient à revoir).

RÉSUMÉS

Le *Dictionnaire du moyen français* est fondé en grande partie sur un vaste corpus informatisé. Le but de cet article est d'examiner de façon concrète l'atout que représente l'exploration d'un tel corpus pour la rédaction d'articles concernant des verbes fréquents et polysémiques (*croire*, *cuidier*, *penser*, *savoir*, *connaître*). L'accent est mis sur deux fonctionnalités offertes par Frantext : l'étude de voisinage et la recherche de cooccurrences, qui permettent au rédacteur d'affiner la description des unités lexicales traitées ; moyennant une phase de tri « manuel », il est possible par exemple d'établir le paradigme des noms figurant en fonction d'objet pour une acception donnée ou encore la fréquence de certaines constructions syntaxiques. Ce sont là des apports

importants pour un dictionnaire qui vise à donner des descriptions explicites des usages – ce qui n'est pas toujours le cas dans les dictionnaires de français médiéval antérieurs.

The *Dictionnaire de moyen français* is mostly based on a large digital corpus. This paper aims at examining how useful it can be to delve into such a corpus, especially when one has to write articles about polysemic verbs as frequent as *croire*, *cuidier*, *penser*, *savoir*, *connaître*. The emphasis is put upon two functionalities offered by Frantext : the study of words surrounding the occurrences of a given word and cooccurrence search. Both enable the researcher to describe, more accurately, the lexical units he is dealing with. After manual selection procedures, it is for instance possible to establish the paradigm of nouns acting as objects for a given meaning of the verb or to assess the frequency of certain syntactic constructions. These are important functionalities, as the dictionary aims at providing explicit information about uses in context, and it was not always the case with previous French dictionaries.

INDEX

Mots-clés : dictionnaire du moyen français, exploitation de corpus, lexicographie, verbes d'opinion, verbes de connaissance

AUTEUR

CORINNE FÉRON

Université du Maine (Le Mans), CELTED (Université de Metz)