

Martin-D. Gleßgen / Claire Vachon (†)

L'étude philologique et scriptologique du *Nouveau Corpus d'Amsterdam*

1. Description philologique du *Nouveau Corpus d'Amsterdam* (NCA)

Le *Nouveau Corpus d'Amsterdam* repose sur le *Corpus d'Amsterdam* établi jadis sous la direction d'Antonij Dees et de Piet Van Reenen. Le balisage grammatical initial a été élargi et enrichi d'informations lexicologiques par les soins d'Achim Stein et de Pierre Kunstmann; nous avons, quant à nous, entrepris une description philologique des *ca* 300 manuscrits et 150 œuvres en question qui rendent compte de la diversité des genres non documentaires de l'ancienne langue française (poésie et prose littéraires, textes religieux, littérature morale, historiographie).

Nous avons opéré à cette fin une révision bibliographique systématique qui poursuivait les trois objectifs suivants:

- (i) compléter les –nombreuses– informations manquantes dans l'ancienne bibliographie (textes ou manuscrits non identifiés, éditions critiques ou diplomatiques);
- (ii) vérifier et compléter la datation et la localisation des manuscrits du *Corpus*;
- (iii) introduire la datation et la localisation des œuvres transmises par les manuscrits en question, élément auparavant absent.

Nous nous sommes basés dans un premier temps sur le *Complément bibliographique* du DEAF, puis, si nécessaire, sur les éditions elles-mêmes. Dans cette optique, l'une d'entre nous a mené à bien de façon systématique les opérations suivantes:

L'identification univoque des manuscrits transcrits dans le *Corpus* et des œuvres correspondantes.

L'indication des codes DEAF des textes et éditions (en créant, si nécessaire, de nouveaux codes, transmis à l'équipe du DEAF).

L'introduction d'un champ «EditionNCA» réunissant les champs préexistants «édition», «titre» et «manuscrit», tout en les complétant; ce champ permet un aperçu rapide sur les éléments bibliographiques essentiels.

L'indication du nombre de feuillets couverts par le texte transcrit et des numéros de feuillets qui lui correspondent dans le manuscrit.

La vérification du passage transcrit pour établir si le texte intégral a été saisi ou si le NCA comporte uniquement un extrait (avec, alors, l'indication du passage en question).

La vérification et l'indication des lieux et dates des compositions et des manuscrits, avec des champs complémentaires précisant la source des informations (DEAF ou une édition donnée).

Cette opération, longue et fastidieuse, a demandé une extrême attention dans le détail. Elle fut commencée par l'un d'entre nous dès 2004, aidé par plusieurs collaborateurs; les premiers résultats ont permis une première synthèse en 2007 (Gouvert / Gleßgen 2007); depuis 2008, C. Vachon a repris nouvellement et complété toutes les informations disponibles, ce qui a permis la synthèse actuelle qui dépasse considérablement la version de 2007.¹ Nous sommes très reconnaissants à notre ami Gilles Roques de nous avoir apporté une importante série de compléments, grâce à sa connaissance unique des textes anciens.

Au final, notre révision a non seulement fortement enrichi la bibliographie initiale mais elle a également permis de corriger une série importante d'erreurs dans la bibliographie de Dees (cotes de manuscrits, dates d'éditions, passages concernés, etc.), voire dans la composition même du corpus (doublets, présence de copies tardives). Voici quelques cas de figure, en dehors des corrections portant sur des erreurs de saisie ou d'inattention dans l'indication du passage ou les cotes des manuscrits:²

Mauvaise identification:

– *mir* (CoincyI10K) correspond au *Miracle de Théophile* de Gautier de Coinci, et non pas aux *Miracles de Notre Dame de Soissons* du même auteur.

Manuscrit erroné:

– *marga* (WaceMargaF) provient du ms. A, et non du ms. M;
 – *sully2* (SermMaurPB) est issu, comme l'indique le DEAF, du ms. Poitiers 97 (271), et non du ms. Poitiers, Bibl. pub. 124;
 – *thebefrag* (ThebesR_D) correspond au ms. Angers BM 22 (26), et non au ms. Angers BM 2.

Erreurs d'édition d'origine:

– l'année de parution correspondant à *enf* (EvEnfB) est 1976, et non 1980;
 – le volume de la revue dont provient *kathe* (SCathJong1F) est R 58, et non R 63.

Fichier présentant un texte fautif:

– le mot de la rime manque systématiquement dans *jongl* (PassJonglGP);
 – le texte d'*amou* (BestAmOctT) est présenté dans l'ordre alphabétique des vers.

Textes à supprimer:

– *fierens* (CartTournaiF*) est un recueil de chartes du 15^e siècle et n'a pas sa place dans un corpus de textes non documentaires;
 – *perb* est inclus dans *percevalb* (PercLo_B);
 – *chret1* et *chret2* contiennent chacun la transcription partielle des deux mêmes œuvres, transmises chaque fois dans un même manuscrit (ms. Paris, BN fr. 1450 pour *chret1* et ms. Paris, BN fr. 794 pour *chret2*); ce qui est grave, c'est que chacune des quatre transcriptions en question est contenue dans d'autres items du *Corpus* (*chret1* est inclus dans *yvf* (YvainL_F) et *perr* (PercLo_R); *chret2* est contenu dans *yvh* (YvainL_H) et *pera* (PercLo_A)).

¹ Prenons comme seul exemple les changements concernant la date du manuscrits, en faisant abstraction des changements purement formels ou «automatiques», comme le remplacement de «1310pm10» par «1313pm13» ou de «1300ca» par «1290pm10» (d'après les règles instaurées par Achim Stein). Depuis 2007, 45 indications concernant la date ont pu être améliorées et précisées; dans 13 cas, le changement ne porte que sur quelques années mais dans 28 cas, il s'agit de corrections majeures: par ex., la date de *ThebesC_D* (= Dees thA) passe de 1400ca à 1200pm10, celle de *AthisH* (= athi) de 1235ca à 1350pm50; par ailleurs, une nouvelle date a été ajoutée (*fierens*: 1489) et trois informations relatives aux dates ont été supprimées car elles n'ont pas été validées (*gar*: 1750pm50).

² Erreur d'un chiffre dans des cotes de manuscrits: *fab4f* et *mous*; erreur de passage: *alia* et *abe*.

2. Perspectives pour l'étude scriptologique

Ces corrections permettent dorénavant une utilisation plus sûre du corpus, tel qu'il est disponible sur le serveur de l'université de Stuttgart.³ Par ailleurs, notre révision philologique a ajouté un nombre important d'informations à la description initiale des textes. À titre d'exemple, il est maintenant possible de prendre en considération, conjointement, les paramètres diasystématiques qui caractérisent les manuscrits et ceux qui sont propres aux textes originaux sur lesquels ils reposent. Cela ajoute une nouvelle dimension interprétative qui a un impact variable selon le domaine de la langue concerné:

- dans les domaines syntaxique et même lexical, les choix diasystématiques des manuscrits suivent fortement le modèle des versions antérieures, ce qui crée le décalage spatio-temporel caractéristique des textes médiévaux tels que nous les connaissons. Cette observation ne met pas en cause la forte variance textuelle due à l'action des copistes et adaptateurs: mais il s'agit là d'interventions motivées essentiellement par des questions de contenu et dans une moindre mesure par un souci d'adaptation linguistique;
- en graphématique et, partiellement, dans les marques morphologiques, les paramètres diasystématiques de l'espace et du prestige social, de même que les changements dus au temps, transforment plus fortement les données de l'original et peuvent mener à une physionomie beaucoup plus proche du dernier scribe que de celle de l'auteur.

Dans un premier temps, il nous a semblé particulièrement intéressant de poursuivre, sur la nouvelle base philologique que nous avons constituée, les analyses scriptologiques d'A. Dees et de P. Van Reenen. Une telle réflexion porte alors sur les éléments de graphématique et de morphologie qui montrent la plus forte intervention des copistes; ces domaines permettent en même temps de tirer le plus grand nombre de conclusions linguistiques sur les caractéristiques des manuscrits individuels.

Etant donné la lourdeur de toute analyse scriptologique de détail, nous étions obligés de restreindre le domaine d'observation à l'intérieur des 300 textes du corpus. Nous avons retenu les textes provenant de l'Est et du Sud-Est du territoire d'oïl (Champagne, Lorraine, Franche-Comté, Bourgogne); cela nous permet notamment de comparer les paramètres scriptologiques des manuscrits littéraires avec ceux des chartes originales, contemporaines, paramètres que nous étudions dans le projet des *Plus anciens documents linguistiques de la France* (Gleßgen 2009, Gleßgen / Kihāi / Videsott 2010).

³ Cf. l'adresse indiquée dans la bibliographie sous *NCA*; le site est accessible gratuitement après inscription à cette même adresse.

3. Etude scriptologique des manuscrits orientaux

3.1. Segmentation géolinguistique du corpus

La restriction du domaine d'observation aux seules quatre régions de l'Est mentionnées suppose tout d'abord l'identification des textes qui se placent dans ce cadre géographique. Cette première opération s'est avérée assez compliquée et nous avons dû commencer notre étude scriptologique par une étape préliminaire relativement lourde; cette première phase sera toutefois pleinement mise à profit pour l'évaluation linguistique ultérieure des manuscrits retenus.

Nous avons commencé par réunir tous les textes qui ont été attribués à l'une des quatre régions de l'Est soit par Dees soit par la tradition philologique, telle qu'elle est répertoriée dans le DEAF ou telle qu'elle s'exprime dans les éditions en question. Cela a fourni un premier groupe de 99 textes (après suppression des deux textes mentionnés, *chret2* et *perb*, cf. *supra* 1.) que nous avons étudiés de plus près d'un point de vue scriptologique. Après avoir cerné les paramètres linguistiques les plus pertinents et avoir dégagé les premières conclusions, nous avons étendu nos interrogations à l'intégralité des textes restants, au nombre de 193 (après suppression des deux autres textes mentionnés, *chret1* et *fierens*). Cet élargissement s'est avéré indispensable pour pouvoir établir la valeur scriptologique des paramètres retenus sur le territoire d'oïl.

Comme nous l'avons dit, le choix des quatre régions retenues a été dicté non pas en premier lieu par des arguments géolinguistiques mais par l'état de nos études sur les chartes françaises originales du XIII^e siècle.⁴ Notre finalité a été toutefois celle d'établir ensuite une segmentation externe et interne, cohérente d'un point de vue géolinguistique et donc basée avant tout sur des critères linguistiques. Dans la segmentation externe, notre étude nous a amené à exclure les zones qui se trouvent dans la périphérie occidentale de notre territoire, proche de l'Île-de-France, puisque les textes n'y présentent pas les principaux paramètres linguistiques que nous avons dégagés comme définitoires de la région géolinguistique de l'Est.

3.2. Les paramètres linguistiques définitoires de l'Est

L'identification des paramètres graphématiques et morphologiques pertinents a été une opération exigeante en termes de temps et de réflexion. Dans un premier temps, nous nous sommes basés sur les paramètres déjà retenus pour nos chartes (cf. Gleßgen 2008, Tinner 2009); malgré l'utilité de ce choix, la nature des manuscrits du NCA, intégralement transmis par copie, induit des types de variation différents de ceux des chartes originales. Certains paramètres sont particulièrement porteurs pour les chartes en raison de leurs choix lexicaux, comme les variations *lettres* ~ *lattes* et *estable* ~ *estau(b)le*, retraçables dans presque chaque document; mais dans les textes du NCA, ces variables paraissent bien entendu de manière moins systématique. Par ailleurs, nous

⁴ Les départements, pour lesquels nous disposons actuellement de l'édition électronique, accessible par internet (cf. Gleßgen 2009) sont les suivants, du Nord-Ouest au Sud-Est: la Marne, la Haute-Marne, la Meuse, la Meurthe-et-Moselle, la Haute-Saône, la Saône-et-Loire, la Nièvre et le Doubs. Tous ces documents ont été balisés avec les paramètres scriptologiques retenus par nous (cf. Gleßgen 2008).

n'avons pas encore pu faire appel à l'apport très précieux de l'analyse codicologique et paléographique des documents, puisque celle-ci suppose d'avoir réuni au préalable la reproduction photographique et la description détaillée de tous les manuscrits du corpus.

Nous avons donc opéré un grand nombre de tests pour pouvoir établir la pertinence des différents paramètres, en nous basant chaque fois sur le nombre des occurrences concernées et sur les pourcentages respectifs des différentes variantes. Les paramètres pertinents doivent répondre à plusieurs exigences (cf. Gleßgen 2008):

- ils doivent être relativement fréquents dans l'ensemble du corpus et dans un nombre important de textes individuels;
- étant donné la dimension du corpus, l'identification et le balisage des paramètres doivent demander un temps défini et ne pas impliquer des vérifications étymologiques trop lourdes;
- la distribution des variables doit répondre à une logique géolinguistique et les variations des pourcentages relevés doivent être significatives.

Voici un bref descriptif de la procédure de base pour le relevé d'une variable (par exemple *-aubl-*) à l'aide du formulaire de recherche mis en ligne par Achim Stein sur le serveur de l'Université de Stuttgart (cf. *NCA*):

- sous *Search words*, inscrire « *.*aubl.** » comme *word or regular expression* (*.** correspond à une séquence quelconque de signes) et activer *statistics on*;
- pour simplifier l'affichage des résultats, mettre *deaf* dans «*or only some attributes*» sous *Line annotation*;
- lancer enfin la recherche en cliquant sur *Start* (en haut à gauche).

Le téléchargement, sur le site du *NCA*, du logiciel *TigerSearch* permet d'enregistrer un sous-ensemble de textes à étudier, ce qui a été nécessaire pour nous; le logiciel permet par ailleurs de compter automatiquement le nombre d'occurrences par texte et d'exporter les résultats en format Excel, ce qui n'est pas possible avec le formulaire de base.

Dans nos sondages, nous avons constamment comparé les renseignements déductibles des différents paramètres afin d'établir lesquels étaient particulièrement porteurs et lesquels n'apportaient aucune information complémentaire par rapport à d'autres paramètres. Cela nous a permis de constituer des hiérarchies relatives que nous ne pourrions pas présenter intégralement en ce lieu mais qui sont parfois surprenantes (cf. les exemples *infra*).

A l'issue d'une importante série de tests, nous avons retenu vingt paramètres pertinents. Parmi eux, six paraissent avec une fréquence relativement élevée dans le corpus global et sont par conséquent potentiellement à haute fréquence dans les textes individuels (la forme purement graphématique *k- vs qu-*, les variables grapho-phonétiques *or ~ our*, *nr ~ ndr* et *abl ~ au(b)l* ainsi que la marque morphologique *le ~ lo(u)*). Par ailleurs, la hiérarchisation des phénomènes a permis de distinguer des paramètres régionalement fortement marqués, comme *lo(u)*; ce sont des paramètres très révélateurs pour une ou plusieurs des quatre régions; leur présence peut également être interprétée en termes diastratiques puisqu'il s'agit de formes d'un moindre prestige socioculturel. D'autres paramètres comme le maintien de *nr* (*genre* pour *gendre*) sont certes caractéristiques par leur fréquence pour l'une de ces quatre régions mais ne

sont pas exclusifs. Enfin, trois des paramètres retenus sont des marques faiblement régionales et correspondent soit à des états chronologiques donnés, comme *or* (contre *our*), soit à des marques à valeur diastratique (*k-* et *w-*).

Nous avons par conséquent regroupé les vingt paramètres dans six catégories distinctes selon leur valeur de marquage et leur fréquence:

(i) Paramètres marqués d'un point de vue régional (et diastratique)

(a) Paramètres à haute fréquence dans le corpus:

= art.m.sg.rég. *lo* ~ *lou* vs *le*

= *aul(e)* ~ *aubl(e)* (avec vélarisation de *-b-* devant *-l*) vs *abl(e)*

= dipht. de /^ʰa/ (lat. *-atu-*, *-ata-*, *-are*): p.p. et inf. en *ei* ~ *eir* (avec dipht. en /ej/) vs *e* ~ *er*

= 6^e pers. en *-unt* vs *-ont* (fut. ou prés. de *estre*, *avoir*, *faire*, *aller* et leurs composés)

(b) Paramètres à faible fréquence dans le corpus:

= *aul* vs *al* (< *-ale*)

= *-a(u)l* vs *-el* (< *-ellu*)

= 3^e pers.prés. avoir: *ai(t)* ~ *at* vs *a*

= pron.pers. tonique *mi* vs *moi* ~ *moy*

= 3^e pers.parf. *estre*: *fu(s)t* ~ *fui(s)(t)* vs *fu*

= démonstratif neutre (< *ecce* + *hōc*): *ceu* vs *ce*

= variantes de *chose*: *-ou-* ~ *-oi-* vs *-o-* / *-z-* ~ *-ss-* ~ *-c-* vs *-s-*

(ii) Paramètres moyennement marqués d'un point de vue régional

(a) Paramètre à haute fréquence dans le corpus:

= *-nr-* vs *-ndr-* (< lat. *n'r*, *prehendere*, qq noms propres)

(b) Paramètres à faible fréquence dans le corpus:

= var. graph. *-aige* vs *-age* (suff. < lat. *-aticu* et rad.: *saige*, *naige*, etc.)

= *-oil* vs *-eil* (subst. *conseil*, *soleil*, *orteil* et *sommeil* uniquement)

= *sein(c)t* vs *sain(c)t* et dérivés (lat. /^ʰa/ devant nasale)

= *noz* vs *nos* ~ *nous*

= *signeur* vs *seigneur* et dérivés

(iii) Paramètres peu marqués d'un point de vue régional

(a) Paramètres à haute fréquence dans le corpus:

= diff. résultats de /^ʰō/ lat.: *or* (*signor*, *lor*) ~ *lor* ~ *cort* ~ *nos* vs *our* etc.

= *ke*, *ki* vs *que*, *qui*

(b) Paramètre à faible fréquence dans le corpus:

= présence de *-w-*, d'origine germanique ou non (fréquence absolue uniquement)

Les paramètres sous (i) sont évidemment les plus décisifs pour l'inclusion ou l'exclusion d'un texte donné dans le cadre géolinguistique de l'Est; mais la cooccurrence d'autres paramètres peut apporter des éléments complémentaires et elle permet, surtout, de mieux décrire par la suite la physionomie des différents textes.

Notre document de synthèse contient un tableau-excel de tous paramètres par texte avec le nombre d'occurrences absolues et les pourcentages. Les degrés de gris dans les colonnes

de pourcentages indiquent des tranches de fréquence différentes, toutes supérieures à 25% (blanc: 0-24%, gris clair 25-49%, gris moyen: 50-74%, gris foncé: 75-100%); cf. comme exemple un segment de ce tableau:

	A	B	C	D	E	F	G	I	J	K	L	M	O	P	Q
1	Signe CEAF	Signe Oeas	ABL	AUBL	AUL	% AUBL	% AUL	LE	LO	LOU	% LO	% LOU	NDR	NR	%NR
2	JPhoatR	abraja	2	37		94,67	0	71			0	0	0	0	0
3	Eles_AB*	ailaa	1			0	0	29			0	0	0	0	0
4	Eles_BB*	ailab						21			0	0	0	0	0
5	Eles_DB*	ailad						28			0	0	0	0	0
6	AmAnD	amila	0			0	0	100			0	0	0	0	0
7	SaintAbb*	arh	74			0	0	42	124	3	73,37	1,75	15	10	36,71
8	ToamAnW	arti	19			0	0	00			0	0	0	0	0
9	JMeunV6yR*	arok	21			0	0	00			0	0	0	0	0
10	BaJocFr_1*	barlaone	22	122		04,72	0	203	1	04	0,32	17,00	90	0	0,20
11	BarrD*	barri	24			0	0	122	1	1	0,01	0,01	22	0	0
12	DialAnD	beatl	20	2	0	0,00	10,10	4	07	2	00,40	0,17	4	0	0,00
13	SBenAn_23	berm2			0	0	100		174	17	01,10	0,00	0	10	00,00
14	BoagCarM*	beurg	0	0		01,04	0		1	00	1,04	00,00	0	0	0,00
15	CalendreS	cales	10			0	0	07	1		1,72	0	7		0
16	SHILC*	carp	44			0	0	34	127	4	70,07	2,42424	24	10	29,41

Sur cette base, nous avons établi une synthèse hiérarchisée des paramètres, regroupés en 18 unités:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	AU(B)L	LOU	NR	NBL	NBL	B	UNT	AUBL	Seint	Cho	soieu	MI	AJ(T)	FUJSTI	KW	EUR	OR	OUR	
2	AU(B)L	x	70,07	70,07	50	14,29	50	70,07	71,43	21,43	14,29	07,14	42,00	00,70	07,14	04,29	7,14	100	00,71
3	LOU	22	x	00	70	12	34	00	20	10	10	00	20	40	40	00	20	100	10
4	NR	13,50	03,00	x	00,70	11,11	22,22	44,44	20,00	10,50	17,20	20,00	22,22	07,04	07,04	70,31	04,07	100	10,70
5	NBL	12,70	03,04	03,04	x	0,40	10,30	40,00	10,10	20,40	20	27,27	21,00	00,10	00,10	00	20,40	100	14,00
6	NBL	22,22	00,07	100	33,33	x	44,44	77,70	22,22	0	00,00	00,00	00,00	00,00	00,07	100	00,00	100	11,11
7	B	00	00	00	40	20	x	70	00	10	20	00	00	00	00	00	00	100	00
8	UNT	00,00	07,44	03,72	03,70	10,20	04,00	x	04,00	10,00	00,00	00,00	00,00	03,40	40,01	70,00	00,00	100	00,00
9	AUBL	00,00	00,42	00,42	00,00	10,00	00,00	70,00	x	10,70	01,00	00,10	00,00	70,00	70,00	00,00	10,70	100	47,07
10	Seint	10,70	00	00,70	07,00	0	00,00	10,70	x	0,00	01,00	10,70	20	00	00	00	00	100	10,00
11	Cho	13,30	00	00,00	73,30	20	00,07	73,30	40	0,07	x	40	00,00	00,00	00	00	00,00	100	00
12	soieu	04,70	00,00	01,30	00,22	10,04	47,00	73,01	02,17	21,74	00,00	x	04,70	00,22	00,07	73,01	21,74	100	00,40
13	MI	27,27	00,00	01,00	04,00	10,04	01,00	00,04	22,70	10,04	22,70	00,00	x	40,01	01,00	72,70	22,70	100	01,00
14	AJ(T)	00,00	72,70	00,01	00,04	10,10	40,00	00,70	40,40	10,10	24,24	40,40	27,27	x	00,04	00,70	10,10	100	07,27
15	FUJSTI	22,22	00,07	00,00	00,00	00,07	00,11	00,00	41,07	22,22	20	44,44	10,44	00,00	x	00,07	07,70	100	10,44
16	KW	13,40	00,70	01,04	40,20	10,40	22,00	40,27	14,00	11,04	17,01	20,07	20,00	04,00	00,00	x	00,04	100	10,40
17	EUR	0,20	41,04	00,00	40,10	0,00	10,20	20,00	0,00	10,10	10,10	10,10	10,10	00,20	70,07	x	100	00,00	
18	OR	13,00	40,00	00,00	04,40	0,01	10,00	40,07	10,01	10,04	14,00	20,77	21,70	00,07	00,04	00,04	00,00	x	17,00
19	OUR	07,70	44,44	00,00	44,44	0,00	00,00	72,22	00	11,11	10,07	00,00	00,00	00	00,00	01,11	44,44	100	x

La lecture de ce dernier tableau demande un peu d'accoutumance; prenons deux exemples:

- la ligne 18 (OR), colonne B (AU(B)L) donne le chiffre 13,86%; ce chiffre indique que 13,86 pourcents des textes qui ont «or» connaissent également les variables «aul» ou «aubl»; la présence du paramètre «or» dans un texte ne suppose donc qu'avec une très faible probabilité la présence du paramètre «au(b)l»;
- la colonne Q (OR) présente le chiffre 100% pour toutes les lignes; cela veut dire que 100% des textes qui ont un des 17 autres paramètres ont également «or»; le paramètre «or» est donc impliqué par la présence de chacun des autres paramètres (= un texte qui connaît «au(b)l» connaît toujours aussi «or» mais l'inverse n'est pas vrai).

Des variables comme «or» ou «nr» (80 à 90% d'inclusion par les autres variables) sont donc endémiques dans les textes orientaux (= elles sont dominantes voire exclusives dans tous les textes orientaux), sans nécessairement avoir une valeur diasystématique très marquée (= elles paraissent également dans des textes non-orientaux).

Le relevé intégral des occurrences et des pourcentages des différentes variables pour les vingt paramètres nous a permis d'obtenir le résultat envisagé: sur cette base, il est possible d'établir une opposition univoque entre les textes à inclure dans les régions de l'Est et les manuscrits exclus; nous avons donc pu opérer un tri sûr pour l'intégralité des 99 textes qui nous intéressent de plus près. Nous avons également pu confirmer, sur cette base, l'exclusion pour près des trois-quarts des 193 textes retenus comme extérieurs à notre région par Dees et la tradition philologique; le dernier quart en revanche présente des profils linguistiques similaires aux textes que nous avons inclus dans notre corpus et devront donc être soumis à un réexamen en ce sens.

Nous présenterons par la suite les premiers résultats de cette distinction pour les 99 textes à l'étude.

3.3. Les textes exclus

Parmi les 99 textes retenus comme potentiellement orientaux, nous avons gardé, après notre analyse, 45 textes, correspondant presque à la moitié de cet ensemble. Parmi les 54 textes exclus de la région géolinguistique de l'Est, se décèlent deux catégories, très différentes:

(i) Des textes rattachés à tort à l'Est, soit par une erreur de la tradition (ce qui est le cas plus rare), soit par une erreur de Dees (cas plus fréquent), soit encore par une erreur des deux:

= Dees place *joinv* (JoinvC) en région parisienne, le DEAFBibl le décrit de composition champenoise (d'après l'éditeur Corbett «un francien teinté de traces champenoises» [37], ce qui correspond sans doute mieux à ce texte destiné à la famille royale) et de manuscrit du Nord-Est. Or, le marquage de l'Est est inexistant: l'on ne relève que 20% de *-nr-*, présent en faible pourcentage sur la plus grande partie du domaine d'oïl, et huit occurrences de *fut*, qui peut caractériser également d'autres régions que l'Est. En revanche, les 99% de formes en *-eur* (*leur*, *seigneur*) font pencher vers une variété centrale bien plus qu'orientale. – La localisation de Dees est donc juste, celle de la tradition erronée.

= Dees place *ailleb* (ElesB_B) en Nièvre / Allier, la tradition donne la composition comme picarde et le manuscrit, dans DEAFBibl comme anglonormand avec des traits bourguignons, chez l'éditeur comme «indeterminate Northern French dialect with a few Anglo-Norman features» (Busby, 4). Ce texte inclut seulement deux de nos paramètres: 100% de *or* et associés, ce qui n'est pas particulièrement significatif, et 50% de *-unt*, ce qui confirme ici la main anglonormande (le paramètre a des valeurs différentes selon les régions), toutefois faiblement présente. – L'erreur est autant de Dees que du DEAF et elle provient éventuellement de la difficulté à gérer deux systèmes coprésents (pic. et agn.).

= *calen* (Calendres) est placé par Dees en Haute-Marne, et est désigné par la tradition comme champenois de composition et de manuscrit. Or, ce constat est infirmé par nos paramètres: 2% de *lo* seulement, une fois *aige* et quatre fois *consoil*, ce qui ne suffit pas à justifier pas une telle unanimité. – La localisation de Dees est erronée, celle de la tradition est juste d'un point de vue purement géographique mais non pas d'un point de vue géolinguistique, comme nous le verrons immédiatement après dans la catégorie (ii) à laquelle ce manuscrit doit être rattaché. Nous reviendrons par ailleurs plus loin sur le cas de ce manuscrit (cf. *infra* 3.4.).

(ii) La deuxième catégorie est composée de textes issus de zones périphériques par rapport à l'Est et/ou à placer géographiquement entre la scripta de l'Île-de-France (ou «francienne» pour utiliser ce terme avec toute la prudence nécessaire) et les zones orientales; selon la distribution, nous avons pu retenir parmi ces textes deux ensembles géolinguistiques principaux correspondant à la périphérie occidentale de l'Yonne et de la Champagne:

= *fetrom* (FetRomF1) [Dees: Yonne; – Tradition: comp. frc., ms. bourg.] comporte de très rares paramètres de l'Est, ni les plus significatifs, ni en grand nombre (1 occ. de *mi* et de *fut*, 2 occ. de *seint*). On ne relève ni *lo*, ni *aubl*, ni *unt*, pourtant caractéristique de la Bourgogne. On se trouve face à une scripta de type «francienne» qui tend vers l'Yonne.

= *fab4e* (AubereeN_transE*) [Dees: Nièvre, Allier; – Tradition: comp. pic., ms. «francien», traits lorr.] contient très peu de paramètres, tous peu marqués: 37% de *-nr-*, 100% de *lor*, 2 occ. de *choze*, 1 occ. de *fut*, de *consoil* et de *w*. Ce texte se situe entre une scripta de type «francienne» et le champenois.

Nous ne pouvons pas présenter, dans ce cadre, tous les résultats pour ces deux groupes ni expliquer plus en détail les raisons de notre choix. Mais nous reviendrons plus loin sur notre catégorisation des paramètres géolinguistiques et nous souhaiterions souligner l'intérêt du deuxième groupe retenu pour la segmentation géolinguistique. En effet, ce groupe montre que la catégorisation géolinguistique traditionnelle qui suit, autant chez Dees que dans la tradition philologique, une répartition selon des régions historiques ou selon les départements actuels, n'est pas pleinement probante pour les régions scriptologiques médiévales. Un texte peut donc provenir de la région administrative ou historique champenoise mais ne pas s'inscrire dans la physionomie linguistique que nous avons pu identifier comme «champenoise».

Il serait facile d'interpréter ce constat comme une preuve de l'influence de l'Île-de-France sur les régions avoisinantes au XIII^e siècle; mais il nous semble plus cohérent de supposer un modèle dans lequel les zones autour de Paris formaient une unité scriptologique relative avec ce noyau urbain représentant une métropole avec déjà 200 000 habitants vers 1300 et avec une concentration extrême de grands lieux d'écriture. Les processus de neutralisation scriptologique étaient alors plus développés dans cette zone «francienne» élargie qu'ailleurs étant donné l'intensité des processus de koïnisation.

Il est certain que la différence entre notre vision et le modèle traditionnel d'un centre directeur parisien peut sembler faible mais elle nous semble importante d'un point de vue épistémologique. Nous n'avons pas pu approfondir en ce lieu cette question, qui pourra être reprise sur la base des études de Paul Videsott sur les chartes de la chancellerie de Paris (cf. Videsott 2010, 2012). Notre vision est toutefois déjà confortée par les résultats de Dumitru Kihai dans son analyse des chartes champenoises (cf. Kihai 2012).⁵

Même si nos résultats sont encore partiels, ils montrent dès à présent la nécessité d'une redéfinition des régions scriptologiques médiévales; celles-ci doivent être détachées des segmentations géo-historiques ou administratives et partiellement même de la réalité

⁵ Il faut, bien évidemment, supposer des zones de transition entre une zone scriptologique et une autre, d'autant plus que la conscience des scribes jouait un rôle bien plus important dans le choix des variables régionales à l'écrit que celle des dialectophones illettrés dont le parler reste toujours proche de la variété natale. Nous verrons par ailleurs que la dimension diastratique des variantes géolinguistiques est pleinement exploitée par les scribes, ce qui introduit une dimension complémentaire dans l'interprétation.

géolinguistique des dialectes modernes voire, peut-être, prémodernes. Les choix des scribes s'inscrivent dans une volonté de marquage conscient et suivent clairement une logique organisée selon des épices et des périphéries de l'écrit. Ces choix opérés par les acteurs médiévaux doivent encore être identifiés par la Science actuelle.

3.4. Caractéristiques des textes inclus

Nous souhaiterions réunir par la suite quelques premiers éléments d'analyse concernant les textes inclus. La petite cinquantaine de textes retenus au total comme pleinement orientaux d'un point de vue linguistique se caractérise par la présence nette d'au moins un paramètre marqué d'un point de vue régional ainsi que de plusieurs paramètres cooccurrents.

Il nous est apparu d'emblée que la présence des différents paramètres est bien plus variable que dans le cas des chartes. Cela s'explique par les différentes étapes de copies mais également par la volonté suprarégionale plus marquée des textes non-documentaires. Nous avons donc distingué la présence des paramètres dans un texte donné selon leur fréquence (haute ou faible), en retenant, toujours après de nombreux tests, cinq tranches de pourcentage ainsi qu'une catégorie de non quantifiée:

- 90 – 100%: exclusif
- 60 – 89%: dominant
- 30 – 59%: moyen
- 11 – 29%: minoritaire
- 1 – 10%: minime
- non quantifié (voire quantification peu significative)

À l'aide de notre catégorisation relativement complexe –vingt paramètres en six catégories, organisés pour chaque texte d'après cinq tranches de pourcentage–, nous avons pu établir une description assez précise de chaque texte.

Notre tableau de synthèse répertorie tous les paramètres linguistiques d'après leurs tranches de pourcentages pour les différents textes orientaux. Le gras indique qu'un paramètre dépasse les 25 occurrences dans un texte. Les textes en gris clair correspondent aux textes exclus, les textes en gris foncé à ceux retenus comme orientaux. Le lien entre les marques régionales et les textes en gris foncé est immédiatement perceptible. Voici le tableau pour les 15 premiers textes:

Signe EFAP	Signe Devis	90-100% exclusif	60-89% dominant	30-59% moyen	11-29% minoritaire	1-10% minime	non quantifié
Préamb.	af %a	MPSL	90-100	AI	98 [1 occ.]	EI	90L, 90S0L
Dees_40*	90a			UNT			
Dees_66*	90a	90+H		UNT			
Dees_100*	90a			98 [100], 98H, 98			90L
AmiensC	90a	90a	98				90L, 98H
BrabantC*	90a	90a	LO [+2% LOa]	UNT, 98		FUT, 98 [1 occ.]	90L, 98H, 90S0L
TrajanV001	90a		UNT	90a	98 [100]		90
Stemv001	90a		90a		98	UNT, 90a	90L, 98H, 90S0L
BrabantP_1*	90a	AUL, 90L, 90S0L, 90S0H	AI, AUL, AUL, 90L		LOa [+0,5% LOa], 98H	EI, 90L [1 occ.]	90L, 98H, 90S0L
StemV0*	90a	PBT	90C	90L, 90S0L	90S	LO, LOU, MIT, 90L	90L, 98H, 90S0L
QuilmesC	90a	LO [+2% LOa], UNT, AI, 90L, 90S0H	98	98	AUL, 90L, EI, 90L	AUL, CHOSSC, CEU, 90L	90L
Dees_10_25*	90a	AI, 90L	EI, 90L, 98		90H	CHOSSC	90L, 98H, 90S0L, 90L
BrabantV0*	90a	LO [+2% LOa], AI, 90L, 90S0H	AUL, AKTO, FUT, 90L, CEU, 90L		AUL, 90L	UNT	90L, 98H, 90S0L
CalenneC5	90a	90+H				LO	90L, 90S0L
StemC_2*	90a	90+H	LO [+2% LOa]	UNT, 98	98 [1 occ.]	FUT, 90S0L [1 occ.]	98H, 90L

Comme nous l'avons déjà indiqué, ce qui nous intéresse, c'est l'apparence du manuscrit tel qu'il se présente au lecteur médiéval ou moderne, sa physionomie générale et, par là, le message diasystématique qu'il veut transmettre; notre but n'est pas d'identifier l'origine individuelle du dernier scribe ni encore les éléments linguistiques qui transparaissent des choix de l'auteur initial ou d'autres scribes intermédiaires. Il est évident qu'il s'agit ici de productions linguistiquement composites; mais nous avons voulu identifier les éléments dominants, correspondant autant à l'histoire du texte qu'à la conscience linguistique du scribe concret du manuscrit.

Les résultats de notre étude permettent non seulement dans de nombreux cas de préciser l'origine du manuscrit par rapport aux indications de Dees ou de la tradition philologique; mais ils introduisent également une nouvelle catégorie interprétative par le degré du marquage diatopique qui se fait l'écho d'une tradition textuelle définie et qui contient en même temps une dimension diastatique voire diaphasique. Le marquage diatopique est naturellement plus présent dans le cas où l'original provient de la même région que la copie, cas de figure favorisé par Dees; mais le genre textuel et la volonté d'une dérégionalisation de prestige jouent, eux-aussi, un rôle important.

Prenons comme exemple deux textes ayant une tradition textuelle semblable mais avec des choix scriptologiques divergents:

Les sermons de Saint Bernard de Clairvaux (éd. Foerster 1886) ont été composés en Lorraine et le manuscrit est également considéré comme lorrain. Il s'agit en effet d'un manuscrit très fortement marqué d'un point de vue régional qui comporte plusieurs paramètres marqués de haute fréquence (*lo*, *aul*, *ei* / *eir*, *unt*) et une forte présence de paramètres de faible fréquence (*at*, *mi*, *fut*, *al*, *ceu*). Voici l'aperçu des formes pertinentes:

hard (SBernAn1F)

(i-a) Paramètres régionalement marqués à haute fréquence:

exclusifs: *lo* 99%, *-aul-* (vs *-abl-*) 100% / dominants: *-ei* 84% ~ *-eir* 60%, *-unt* 65%

(i-b) Paramètres régionalement marqués à faible fréquence:

exclusifs: *at* 100%, *mi* 100%, *fut* 100%, *-al* (vs *-el*) 100% (1 occ.) / dominants: *ceu* 90%

(ii) Paramètres moyennement marqués d'un point de vue régional

-nr- 71% / 6 occ. de *-aige*, 8 occ. de *consoil*, 13 occ. de *signeur*

(iii) Paramètres peu marqués d'un point de vue régional

-or et associés 100%, *ke* 94%

En revanche, le *Tournoi de Chauvency* de Jacques Bretel (éd. Delbouille 1932), dont la composition et le manuscrit se placent également en Lorraine montre une physionomie plus étonnante: le marquage oriental ressort dans des fréquences relatives faibles ou minimales (0,5% *lou*, 8% *ei*), et seulement les paramètres moyennement ou peu marqués d'un point de vue régional (*nr*, *or*) sont fortement présents dans ce manuscrit:

chauvency (BretTournd)

(i-a) Paramètres régionalement marqués à haute fréquence:

minoritaire: *mi* 25% / minimales: *lou* 0,5%, *-ei* 8%, *-unt* 2%

(i-b) Paramètres régionalement marqués à faible fréquence:

minimes: *ai* 1%, *fut* 5%, *ceu* 1%

(ii) Paramètres moyennement marqués d'un point de vue régional

moyen: *-nr-* 44% / non quantifiés: 23 occ. de *-aige*, 24 occ. de *signeur* et 4 occ. de *consoil*

(iii) Paramètres peu marqués d'un point de vue régional

exclusif: *-or* 97% / moyen: *jor* 59% / non quantifié: 45 occ. de *w*

Des décalages dans le marquage régional peuvent même faire surface dans la comparaison entre deux parties d'une même oeuvre –d'origine champenoise– et provenant d'un même manuscrit (de Lorraine): c'est le cas des deux parties de la Genèse d'Evrat contenues dans le ms. Paris BN fr. 12456; elles sont en effet toutes les deux clairement marquées mais se distinguent dans le résultat de /'a/ en syllabe libre (*ei* dans la première partie *e* dans la seconde) ainsi que dans la présence de *w* (presque absent de la première et très fréquent dans la seconde); voici l'aperçu détaillé des variables:

evratC2 (EvratGen)

(i-a) Paramètres régionalement marqués à haute fréquence:

moyens: *lo* 43%, *-eir* 54%, *-unt* 38% / minime: *-ei* 4%

(i-b) Paramètres régionalement marqués à faible fréquence:

minimes: *at* 3%, *fut* 6%

(ii) Paramètres moyennement marqués d'un point de vue régional

moyens: *-nr-* 59%, *ke* 58%

(iii) Paramètres peu marqués d'un point de vue régional

exclusifs: *-or* et associés 95% et plus / non quantifié: 1 occ. de *w*

evrat1 (EvratGen)

(i-a) Paramètres régionalement marqués à haute fréquence:

moyens: *lo* 56%, *-unt* 52%

(i-b) Paramètres régionalement marqués à faible fréquence:

minimes: *at* 2%, *fut* 2%

(ii) Paramètres moyennement marqués d'un point de vue régional

moyen: *ke* 56% / minoritaire: *-nr-* 27%

(iii) Paramètres peu marqués d'un point de vue régional

exclusifs: *or* et associés 100% / non quantifié: 116 occ. de *w*

L'interprétation des différentes formes de variation dans le marquage régional demandera un retour à chacun des manuscrits et chacune des œuvres en particulier. Comme nous l'a fait remarquer Gilles Roques lors de la discussion à Valence, il faut prendre en considération le cas des scribes bien connus comme celui de Guiot, responsable pour la version de *Calendres* que nous avons exclue des textes de l'Est (*supra* 3.3. n° 1); d'un point de vue physionomique, Guiot ne semble donc pas suivre une scripta champenoise mais celle de l'Île-de-France, ce qui mérite toute attention. Par ailleurs, il faut prendre en considération la chronologie: toujours parmi les exclus, le manuscrit de la *Vie de saint Louis* de Joinville date d'environ 1335 et se place donc dans la partie la plus tardive de notre corpus; il s'agira alors de vérifier si ou en quelle mesure la scripta de l'Île-de-France connaît réellement une plus forte présence au XIV^e siècle qu'au début du XIII^e.

Nous avons déjà pu constater que, dans le domaine des genres textuels, les textes religieux de notre corpus connaissent un marquage géolinguistique plus fort que les textes profanes. Cela est partiellement dû à leur mode de transmission, souvent plus local, mais l'on peut y voir également un effet lié à leur diffusion et à leur statut dans la Société.

4. Perspectives

À l'heure d'aujourd'hui, nous avons décrit de la sorte en détail les 99 textes retenus comme potentiellement orientaux, en précisant dans de nombreux cas leur origine au-delà des indications actuellement disponibles. Comme premier résultat de cette description ressort notamment la forte variation dans le degré du marquage géolinguistique entre différents manuscrits provenant d'une même région. Cet élément n'a jamais été suffisamment pris en considération par la scriptologie ni encore par les descriptions linguistiques individuelles des manuscrits.

Notre étude n'est pas terminée; nous devons encore établir de plus près les équivalences et décalages entre les manuscrits non-documentaires et les chartes des mêmes régions. Nous devons également mener à bien la différenciation interne des 45 manuscrits réellement orientaux sur la base des constats établis jusqu'ici et préciser la physionomie des textes rattachable à la scripta de l'Île-de-France. Par ailleurs, nos résultats nous mèneront à reprendre chaque cas de figure en détail pour permettre une interprétation individuelle des textes à l'étude. Mais nous avons pu identifier d'ores et déjà l'importance des dimensions diastratique et diaphasique dans le marquage régiolectal des manuscrits non-documentaires au Moyen-Âge.

Eléments bibliographiques

- DEAFbibl = Möhren, Frankwalt (2007): *Dictionnaire étymologique de l'Ancien Français – Complément bibliographique*. Tübingen: Niemeyer.
- Dees, Anthonij (1987): *Atlas des formes linguistiques des textes littéraires de l'ancien français*. Tübingen: Niemeyer.
- Gleßgen, Martin-D. (2008): *Les lieux d'écriture dans les chartes lorraines du XIII^e siècle*. In: *RLiR* 72, 413-540.
- (2009): *Les plus anciens documents linguistiques de la France. Édition électronique*, Collection fondée par Jacques Monfrin, poursuivie par M.-D.G., en collaboration avec Françoise Viellard et Olivier Guyotjeannin, en partenariat avec Paul Videsott. <www.mediaevistik.uzh.ch/docling>
- / Gouvert, Xavier (2007): *La base textuelle du Nouveau corpus d'Amsterdam: ancrage diasystématique et évaluation philologique*. In: Kunstmann / Stein, 51-84.
- / Kihai, Dumitru / Videsott, Paul (2010) [2011]: *L'élaboration philologique et linguistique des «Plus anciens documents linguistiques de la France, Édition électronique»*. *Trois études*. In: *Bibliothèque de l'École des Chartes* 168, 5-94.
- Kihai, Dumitru (sous presse): *Le scribe bilingue dans les productions documentaires oïliques: étude d'un cas concret*. In: *Actes del XXVI Congrés de Lingüística i Filologia Romàniques*. València 2010. Walter de

Gruyter.

Kunstmann, Pierre / Stein, Achim (édd.) (2007): *Le Nouveau Corpus d'Amsterdam* (Actes de l'atelier de Lauterbad, 23-26 février 2006). Stuttgart: Steiner.

NCA = Stein, Achim / Kunstmann, Pierre / Gleßgen, Martin-D. (édd.) (2010): *Nouveau Corpus d'Amsterdam*. Corpus informatique de textes littéraires d'ancien français (ca 1150-1350), établi par Anthonij Dees (Amsterdam 1987), remanié par A. S., P. K. et M.-D. G. Stuttgart: Institut für Linguistik / Romanistik, version 2-2. <<http://www.uni-stuttgart.de/lingrom/stein/corpus/#nca>>

NCAbibl = *Bibliographie du Nouveau Corpus d'Amsterdam*, établie par A. Dees et P. van Reenen (Amsterdam 1987), remaniée intégralement par M.-D. Glessgen et Claire Vachon, Stuttgart: Institut für Linguistik/Romanistik, version 2-2. <<http://julienas.philosophie.uni-stuttgart.de/nca/version2/download/amslitbib.html>>

Tinner, Sarah (2009): *Les lieux d'écriture dans les chartes franc-comtoises du XIII^e siècle*. Mémoire de maîtrise, Zurich.

Videsott, Paul (2010) [2011]: *À propos du plus ancien document en français de la chancellerie royale capétienne*. In: *BEC* 168, 61-81.

— (sous presse): *Quand et avec qui les rois de France ont-ils commencé à écrire en français?* In: *Actes del XXVI^e Congrès de Lingüística i Filologia Romàniques*. València 2010. Walter de Gruyter.

— (en préparation): *Le «Corpus des actes royaux en français du XIII^e siècle»*.