



HAL
open science

Les prépositions en, dans et dedans au XVIe s. Approche statistique et combinatoire.

Denis Vigier

► **To cite this version:**

Denis Vigier. Les prépositions en, dans et dedans au XVIe s. Approche statistique et combinatoire..
Le Français Moderne - Revue de linguistique Française, 2015, 2. halshs-01276247

HAL Id: halshs-01276247

<https://shs.hal.science/halshs-01276247v1>

Submitted on 10 Oct 2017

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.

***En, dans, dedans* prépositions au XVI^e siècle. Approche statistique et combinatoire.**

Denis Vigier

Si la concurrence entre les prépositions *en* et *dans* en français moderne a fait l'objet d'études nombreuses¹ depuis Guillaume (1919, 1975), cette même concurrence envisagée dans la langue du XVI^e s. a en revanche assez peu retenu l'attention des linguistes. A fortiori la triade *en, dans, dedans* pour cette même période. On ne dispose guère que des travaux de Darmesteter (1885) et de Gougenheim (1945, 1950, 1951), dont les principaux résultats ont été repris et confirmés dans des études récentes (e.a. Sancier-Château (1995), de Mulder (2008), Fagard et Sarda (2009)). Darmesteter - se fondant sur le dépouillement d'œuvres de Rabelais, de Du Bellay, de Scève et de Ronsard - affirmait qu'avant 1550 *dans* est extrêmement rare. « Avec Ronsard, ajoutait-il, *tout change* ». Gougenheim (1951), prolongeant le travail de son devancier, s'est quant à lui intéressé aux « *diverses combinaisons syntaxiques dans lesquelles pouvaient être incorporées les prépositions en et dans* » dans une partie de l'œuvre de Ronsard.

Ces études de détail, extrêmement précieuses, étaient cependant d'emblée limitées par le caractère manuel des dépouillements sur lesquels elles se fondaient. Or on peut aujourd'hui interroger de vastes bases de données textuelles structurées et annotées et, grâce à des outils informatiques mobilisant des calculs statistiques, observer finement les préférences combinatoires des prépositions et leur évolution dans le temps². Les travaux de Blumenthal (2008, 2011, 2012) ont montré que de telles observations permettent en particulier i) d'identifier des affinités sémantiques « profondes » qui s'établissent entre les prépositions et leurs régimes, autorisant du même coup à affiner l'identité sémantique de celles-ci ; ii) d'observer l'émergence et le déclin, en diachronie, de structures figées encore mal connues aujourd'hui.

Le présent article³ propose une étude statistique et combinatoire des prépositions *en, dans* et *dedans* au XVI^e siècle. Nous examinerons en particulier si ces trois prépositions peuvent être distinguées à partir de certaines combinaisons « préférentielles » (i.e hautement significatives sur le plan statistique) dans lesquelles elles entrent. Les corpus sur lesquels nous nous appuyons sont d'une part un ensemble de textes littéraires du XVI^e siècle qui réunit des œuvres relevant de plusieurs genres discursifs, d'autre part un ensemble d'extraits tirés de la base Frantext non catégorisée (période 1500-1599, tous genres). L'outil utilisé est la plateforme TXM développée à l'ENS de Lyon par S. Heiden et son équipe⁴.

¹ Pour une revue récente de ces études, cf. Vigier (2013)

² Le présent article s'inscrit parmi les publications du projet franco-allemand *PRESTO* (*L'évolution du système PRÉpositionnel du français : études STATistique et textOMétrie*) <http://presto.ens-lyon.fr> financé par l'ANR et la DFG (2013-2016), co-porté par P. Blumenthal (Univ. Köln) et l'auteur du présent article.

³ Je remercie L. Bertrand (CESR), M. Clément (Univ. Lyon2), M.-L. Demonet (Univ. Tours), B. Fagard (CNRS-LaTTICe), C. Guillot (ENS Lyon), S. Heiden (ENS Lyon), A. Lavrentiev (CNRS-ICAR), B. Pincemin (CNRS-ICAR) et S. Prévost (CNRS-LaTTICe) pour leur aide précieuse et leurs remarques. Les erreurs sont miennes.

⁴ <http://textometrie.ens-lyon.fr> et Heiden (2010). La version 0.6. de la plateforme été utilisée pour cette étude

1. *En, dans, dedans* au XVI^e siècle

En placé devant *le, les* se contracte dès le Moyen Âge pour donner les formes amalgamées *el, es*. En revanche, il ne se produit aucune contraction devant *la, l'* ou les autres déterminants.

Au XVI^e s., la forme amalgamée *el* (*en + le*) évolue rapidement vers *ou* et *on* qu'on trouve par ex. chez Rabelais. Ces deux formes deviennent progressivement archaïques et disparaissent alors que le pluriel *es* (*en + les*), quoique affecté par le déclin des autres formes amalgamées, se prolonge jusqu'à la fin du siècle (Brunot, 1906 : 277 ; Gougenheim, 1951, 1974 : 181).

Parallèlement, dès le commencement du XVI^e s., *el* puis *ou, on* sont progressivement supplantés par *dedans* (préposition) et par *au* – amalgame de *à* et de *le* ; de même *es* par *aux* (Gougenheim et Brunot, *ibid*). On observe en outre fréquemment dans les textes l'alternance *es* et *aus*

- (1) *non seulement **aux** hommes, mais aussi ès bestes* (Bouch., *Ser*, II, p 3 ; cité par Brunot, *ibid*, 278)

et le flottement est tel entre ces deux formes, précise Gougenheim, qu'on trouve chez Rabelais *es* employé là où *aux* serait attendu :

- (2) *Pantagruel dist **es** invitez* (III, 36) (Voir aussi Brunot, *ibid*.)

La substitution de *au* et *aus* à *el, ou, on* et *es* ne pouvait être durable, du fait en particulier des ambiguïtés que cette situation engendrait. Ainsi dans cette phrase de l'*Heptameron* (citée par Gougenheim 1951, 1974 : 183):

- (3) *Amour ne change point le cueur, mais le monstre tel qu'il est, fol **aux** folles et saige **aux** saiges*

Faut-il interpréter les SP *aux folles* et *aux saiges* comme des datifs régis par le verbe *monstre* (préposition *à*) ou bien comme équivalant à *en + les*, d'où le sens : *fou chez les fous, sage chez les sages*?

Comme l'écrit Gougenheim (*ibid*), *le remède durable vint d'un autre côté*, celui de *dans*, ajoutant après Darmesteter (1890) (repris aussi par F. Brunot) que ce n'est qu'avec Ronsard que *dans* est véritablement introduit dans la langue littéraire⁵. Auparavant, dans les textes littéraires du moins, les emplois de *dans* demeurent rares. Parallèlement à cette évolution qui permet à *dans* de s'imposer progressivement à partir de la seconde moitié du XVI^e siècle, le mot *dedans* se maintient dans son emploi prépositionnel⁶ au-delà du XVI^e siècle (Haase, 1969 : 338). De même pour les emplois de *en + la, l'*.

⁵ Gougenheim précise que Ronsard, dans les corrections successives qu'il apporte à ses éditions à partir de 1567, tend significativement à réduire ses emplois de *dans* (sauf devant *le* et *les*), *comme s'il se repentait de sa hardiesse*. Voir aussi L.Terreux (1968).

⁶ C'est-à-dire suivi d'un régime nominal. Nous considérons ici comme non prépositionnels (mais adverbiaux) les emplois de *dedans* non suivi d'un régime. Par ex. *la barque (...) mais aussi deux matelotz qui estoient dedans pour la garder* (Lery, 1578).

2. Présentation des corpus

2.1. Corpus Epistemon

Les textes de ce corpus sont issus de la base Epistemon des BVH (« Bibliothèques Virtuelles Humanistes » (<http://www.bvh.univ-tours.fr/>)). Il s'agit de 44 textes⁷ regroupant 25 auteurs et près de deux millions de mots. 41 textes appartiennent au XVI^e siècle (les dates d'écriture vont de 1529 à 1587), et 3 au début du XVII^e siècle (1600, 1610, 1616). C'est pourquoi nous avons construit un sous-corpus ne réunissant que les textes du XVI^e siècle (désormais : CEp^{XVI}).

Même si notre corpus est important et répond de façon satisfaisante aux conditions de *pertinence*, de *cohérence*, de *régularité* et de *volume* (B. Pincemin, 1999 : 415-427 ; Guillot, Heiden et Lavrentiev, 2007), il ne prétend cependant pas à une parfaite représentativité. C'est pourquoi les résultats présentés *infra* sont à considérer avec une certaine prudence et devront être éprouvés sur un corpus plus vaste⁸.

Les tableaux 1 et 2 présentent deux partitions opérées sur notre corpus et auxquelles nous recourons à plusieurs reprises dans cette étude. La première a été effectuée sur sa totalité (CEp) ; la seconde sur le corpus restreint CEp^{XVI}.

Années	Total	20-29	30-39	40-49	50-59	60-69	70-79	80-89	1600+
Nbre textes	44	1	4	8	13	10	3	2	3
Nbre mots	1974747	82 473	222 941	214 294	348 706	234 221	218 812	84 573	568 727

Tableau 1 : Partition « Décennies » (corpus CEp)

Auteurs	Tory	Marot	Trad. uc. Castiglione	Rabelais	Scève	Du Bellay	Ronsard	Labé	Léry	Autres auteurs
Nbre Mots	82 473	71 014	145 264	128 598	33 467	38 879	84 597	37 347	117 201	1 170 366

Tableau 2 : Partition « Auteurs » (corpus CEp^{XVI})

2.2. Corpus Frantext

Nous avons aussi importé dans la plateforme TXM deux autres corpus réunissant les extraits obtenus sur la base Frantext (période : 1500-1599, tous genres) pour les requêtes [*dans/dens*] (désormais, corpus CFr^{DANS}) et [*dedans/dedens*] (désormais, corpus CFr^{DEDANS}). Voici la partition⁹ « Décennie » opérée sur ce corpus (désormais corpus CFr, période 1520-99):

⁷ On trouvera sur le site des BVH la liste complète de ces textes avec, pour chacun d'eux, une fiche associée (auteur, date d'édition, imprimeur, librairie, ...).

⁸ Incluant certains auteurs absents de cette version du corpus, dont par ex. Montaigne.

⁹ Pour des raisons techniques que nous ne pouvons pas détailler ici, une partition « Auteurs » n'est actuellement pas possible sur le corpus CFr.

	Total	1500-09	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
Nbre Textes	152	4	7	4	12	21	24	21	24	29	6
Nbre mots	5.785.944	83.463	216.531	184.202	220.503	409.029	903.774	1.325.621	809.493	986.382	646.946

Tableau 3 : Partition « Décennies » (corpus CFr)

Aucune requête n'a été effectuée pour *en* sur Frantext car, dans l'état actuel de cette base, il est impossible de distinguer la préposition du pronom pour cette période.

3. Graphies

Pour la préposition *dans*, deux graphies ont été identifiées : {*dans*, *dens*}¹⁰. Le problème posé par la seconde graphie (peu répandue : 15 occ. dans CEp^{XVI}, 26 occ. dans CFr^{DANS}) est qu'elle se confond avec le pluriel du nom *dent*, beaucoup plus répandu. En l'absence de lemmatisation de nos corpus (cf. *infra*), nous n'avons pas toujours pu inclure cette graphie dans nos calculs. Dans (4), la première occurrence correspond à la préposition, la seconde au nom :

- (4) *Et qui voit ses cheveux encor,
Voit dens Pactole le tresor
Dequoy ses sablons il redore.
(...) Et qui voit ses dens en riant
Voit des terres de l'Orient
Meinte perlette desrobee.
(Labé, Œuvres, 1555)*

Dedans présente aussi deux graphies : {*dedans*, *dedens*}. Désormais (et sauf mention explicite), nous désignerons par les mots *dans* et *dedans* les lemmes et non les mots graphiques.

Pour *en*, il faut enfin envisager le cas des amalgames. Seul CEp^{XVI} peut être ici pris en compte (cf. *supra* § 2.2.). Pour *el* on trouve 6 occurrences ; pour *ou*, aucune occurrence (le mot possède toujours le statut de conjonction). La recherche de l'amalgame *on* pose le problème du pronom : l'absence actuelle d'annotations morphosyntaxiques (cf. *infra*) empêche de sélectionner automatiquement les occurrences relevant de l'une ou de l'autre catégorie. Un balayage rapide de la concordance semble cependant confirmer le point de vue de Gougenheim (1951, 1974 : 181) selon qui cette forme se trouve essentiellement chez Rabelais. Concernant enfin l'amalgame résultant de *en+les*, on trouve trois variantes graphiques (avec ou sans majuscules) : {*es*, *ès*, *és*}.

4. Annotation morphosyntaxique d'un corpus XVI^e siècle par *TreeTagger* dans TXM : état actuel

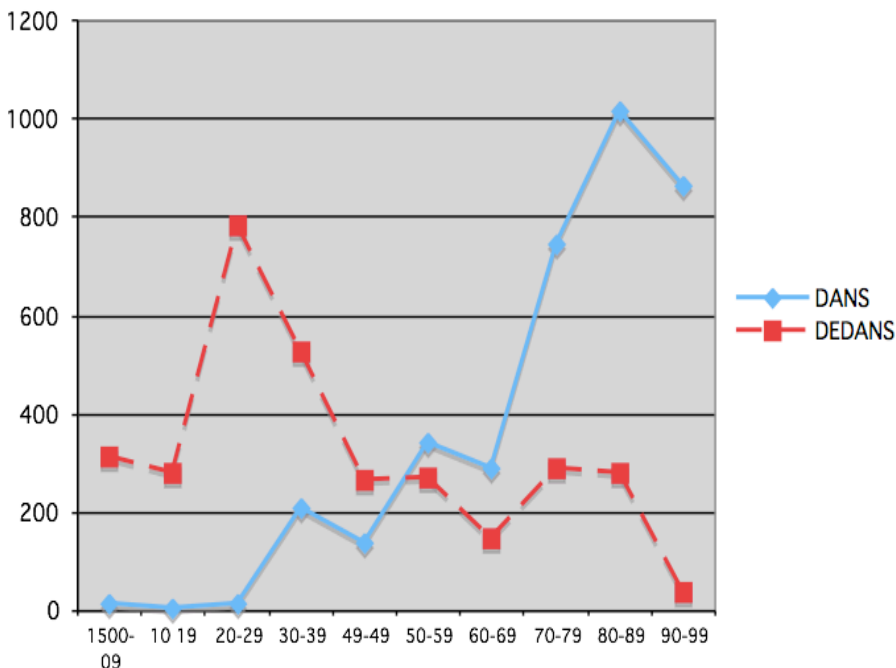
La faible qualité de l'étiquetage et de la lemmatisation de nos corpus ne nous permet pas actuellement de sélectionner une propriété morphosyntaxique ou un

¹⁰ Nous ferons désormais figurer entre accolades les variantes graphiques d'un même mot.

lemme lors de nos requêtes en CQL¹¹ dans TXM. Cette situation n'est pas due à l'outil *TreeTagger* utilisé pour l'étiquetage mais à l'usage que nous en avons fait (le modèle utilisé pour ce corpus étant celui employé pour étiqueter le français moderne). Nous nous sommes donc appuyé sur des requêtes utilisant la forme des mots. Il va de soi qu'à terme, l'étiquetage et la lemmatisation de ces ressources textuelles du XVI^e siècle sont incontournables¹².

5. Evolution des emplois de *dans* et *dedans* au XVI^e siècle

Le graphique 1 permet de visualiser les courbes d'évolution des fréquences relatives (nombre d'occurrences par millions de mots) de *dans* et *dedans* au XVI^e siècle. Le calcul a été accompli sur la partition « Décennies » de Frantext. Le nombre significativement plus élevé d'occurrences de ces deux prépositions dans ce corpus¹³ et leur répartition sur l'ensemble des tranches temporelles envisagées (cf. tableau 3) permet d'approcher avec un peu plus d'assurance (que dans le corpus Epistemon) l'évolution quantitative des emplois de *dans* et *dedans* au XVI^e siècle.



Graphique 1 : courbes des fréquences relatives pour *dans* et *dedans* au XVI^e siècle, corpus CFr, partition « Décennies »

¹¹ *Structured Query Language*, ou *langage de requêtes structuré*.

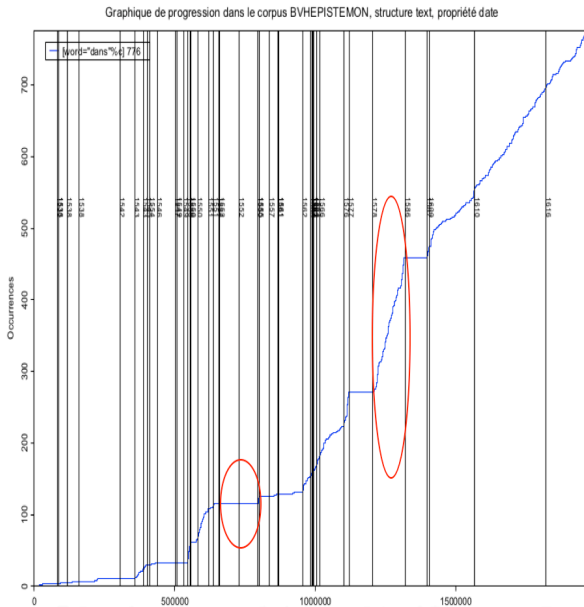
¹² Un autre des objectifs du projet franco-allemand PRESTO vise à développer un lemmatiseur pour le français de cette période.

¹³ CFr^{DANS} présente 6,2 fois plus d'occ. de *dans* que CEp^{XVI} (2989/483), CFr^{DEDANS} 2,2 fois plus d'occurrences de *dedans* que CEp^{XVI} (1345/609).

On observe un accroissement spectaculaire de la fréquence relative de *dans* au cours du siècle tandis que celle de *dedans* préposition décroît globalement durant la même période, les deux courbes se croisant dans la décennie 1550-1559. Ces valeurs nuancent quelque peu les conclusions de Darmesteter (*ibid*, 11-14) et Gougenheim (1945 : 66 ; 1951, 1974 : 183) à propos de *dans*, selon qui cette préposition *est extrêmement rare avant 1550*. Les fréquences observées signalent plutôt un accroissement progressif de *dans* au cours du siècle avec une accélération très forte à partir de la décennie 1560-1569. Les données avancées par Fagard & Sarda (*ibid* : 227), quoique moins précises, corroborent les nôtres.

Quelle part chaque auteur¹⁴ prend-il à l'évolution globalement croissante de l'emploi de *dans* au cours du siècle ? C'est là une question qui avait retenu l'attention de Darmesteter (*ibid*) et de Gougenheim (1945 ; 1951, 1974) mais leur dépouillement manuel des textes limitait *de facto* considérablement le champ de leurs observations. L'application de nos outils nous permet d'aller plus loin.

Le graphique 2 construit par TXM pour la préposition *dans* (corpus Epistemon intégral CEp, textes de 1529 à 1616) permet de visualiser l'accroissement de sa fréquence absolue au sein des textes du corpus rangés par ordre chronologique croissant.

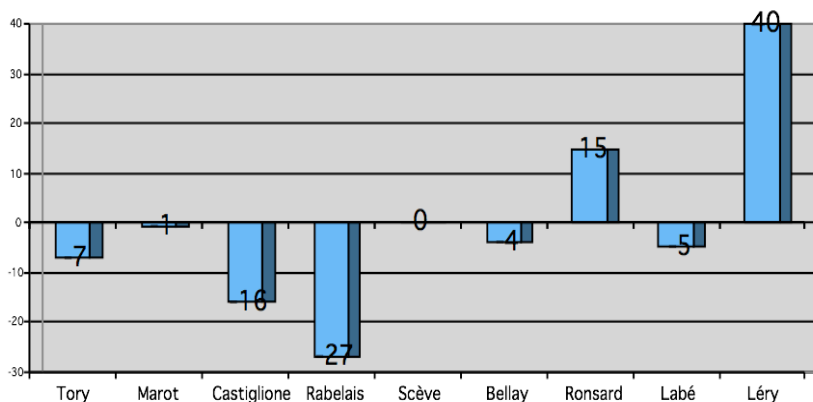


Graphique 2 : Progression de « dans » dans le corpus Epistemon (CEp), valeur de propriété : « date »

¹⁴ Pour des raisons de qualité des textes (qualité des numérisations accomplies, vérification de la propreté des données textuelles, qualité de la conversion en format XML-TEI, ...) et de droits d'éditeurs, mais aussi pour des raisons techniques d'import dans TXM, nous n'avons retenu dans la suite de cette section que le corpus Epistemon. Les résultats présentés concernent donc uniquement les textes littéraires.

Les traits verticaux sur lesquels figurent des dates¹⁵ signalent les années d'édition des textes du corpus, classés par ordre chronologique croissant sur l'axe des abscisses. La distance variable entre deux traits est proportionnée à la taille de chacun des textes. Les pentes raides de la courbe signalent une augmentation forte du nombre d'occurrences dans un texte donné, les pentes douces ou les plateaux, un faible (voire inexistant) accroissement. L'intérêt majeur de cette visualisation est de pouvoir contraster entre eux les textes du corpus (sur le critère de l'emploi de *dans*), et de favoriser l'identification des plus « remarquables » à cet égard. Ainsi avons-nous tracé deux ellipses qui signalent d'une part trois textes de Rabelais édités en 1552 (*Briefve declaration, le Quart Livre, Le Tiers Livre*), d'autre part le texte de Léry *Histoire d'un voyage fait en la terre du Bresil* (1578). Tandis que les premiers ne font (presque) aucun usage de *dans* (une seule occurrence recensée : cf. *infra*), le second en fait un usage contrastivement très soutenu.

On peut se faire une idée encore plus précise de l'attitude de certains auteurs vis-à-vis de l'emploi de *dans* en procédant au calcul de l'« indice de spécificité » (au sens statistique) de *dans* pour chaque tranche temporelle, en tenant compte des variations de taille des parties. Le graphique 3 présente une sélection des valeurs obtenues par certains auteurs figurant dans CEP^{XVI}.



Graphique 3 : spécificité de « *dans* » pour la partition « Auteurs » du corpus Epistemon (CEP^{XVI})

Les signes négatif ou positif affectés aux valeurs numériques signalent respectivement une sous-représentation ou une sur-représentation de l'item comptabilisé. Quant aux valeurs numériques elles-mêmes, elles s'interprètent comme suit : pour Rabelais par ex., la valeur absolue de l'indice calculé (27) indique qu'il y a une probabilité¹⁶ de 10^{-27} que *dans* apparaisse avec une fréquence aussi faible dans l'ensemble de son œuvre. Pour Ronsard, la probabilité que cette préposition apparaisse avec une fréquence aussi forte dans son œuvre est de 10^{15} . Le même

¹⁵ Malheureusement peu visibles ici.

¹⁶ La probabilité calculée porte sur un seuil de fréquence.

raisonnement est à étendre aux autres valeurs. Pour Léry, l'indice de spécificité est si fort qu'il n'a pas pu être calculé (valeur « 40 » de convention).

Les positions symétriques occupées par Ronsard et Rabelais confirment spectaculairement les observations faites par Darmesteter, Brunot et Gougenheim : « *Darmesteter notait (...) que dans (...) ne se trouve point chez Rabelais (...). Avec Ronsard, tout change* » (Gougenheim, 1950, 1970 : 66-67). De fait, chez les auteurs ayant commencé à publier avant 1550 (de Tory à Du Bellay compris), *dans* est nettement sous-représenté dans leur œuvre, Rabelais occupant à cet égard la première place. Peut-on faire de cette absence notoire de *dans* chez Rabelais un trait caractéristique d'une écriture archaïsante ? Ce serait s'aventurer. Mais le fait qu'il n'a jamais inséré dans ses textes, lors des corrections qu'il leur apportait au fil des éditions, une préposition qui commençait à diffuser en français, est bien un signe de cette « *censure à l'antique* » qu'il revendique (Huchon, 1981), d'une volonté *a posteriori* archaïsante.

Il apparaît en outre que contrairement à ce qu'affirme Gougenheim (« [*dans*] *ne se trouve pas chez Rabelais* », 1950, 1970 : 66), il existe une (seule) occurrence de *dans* chez Rabelais, dans un passage de *Gargantua* (1542).

- (5) *Et plus soubdain ne doit estre estimé
Le mouvement que feit Inarimé
Quand Tiphoeus si fort se despita,
Que dens¹⁷ la mer les montz precipita.*

Hapax à première vue étrange... mais qui s'éclaire lorsqu'on retourne au texte. Cet emploi exceptionnel apparaît en effet au chapitre LVIII de l'édition de 1542 du *Gargantua* « *Enigme en prophétie* » emprunté à Mellin de Saint Gelais. Intéressante observation à verser au dossier à charge contre la thèse soutenue par R.L. Frautschi (*French Studies*, XVII, 1963, 331-340) pour qui l'énigme ne serait pas de Saint Gelais mais de Rabelais lui-même (cf. note 5, p 306, édition Droz 1970 de *Gargantua*). Il est clair que cette occurrence seule d'un mot par ailleurs absent de tout le corpus rabelaisien (y compris le corpus douteux du *Cinquième Livre*) suffit à invalider la thèse de l'écriture de ces vers par Rabelais.

Un autre élément remarquable dans le graphique 3 est la position éminente qu'occupe le texte de Léry *Histoire d'un voyage fait en la terre du Bresil* (1578), texte que nous avons déjà remarqué dans l'analyse du graphique 2. Si Ronsard fut bien - si l'on en croit la thèse défendue par Darmesteter et Gougenheim relativement à l'emploi de *dans* dans les textes littéraires - un innovateur en son temps, Léry le fut-il aussi à la fin du siècle ? Pour s'en assurer, il faudrait entrer dans le détail de ce texte de relation de voyage afin de déterminer si la sur-représentativité de *dans* y est simplement due à une nécessité souvent renouvelée de localiser spatialement les événements racontés¹⁸ (dimension du genre discursif) ou bien si l'on observe aussi chez cet auteur des innovations, par exemple dans la combinatoire morphosyntaxique de *dans*, comparativement à celle dont use Ronsard notamment. On signalera pour finir les positions intermédiaires qu'occupent Scève et Marot, et dans une moindre

¹⁷ Dans l'édition princeps, ce *dens* était graphié *dans*. Indice selon lequel Rabelais aurait opté dans la seconde édition de son texte pour une graphie archaïsante (ce commentaire m'a été fait par Michèle Clément, spécialiste du XVI^e s. à l'Université Lyon 2)

¹⁸ *dans* étant, à cette époque, nettement plus utilisé que *dedans* (cf. graphique 1)

mesure Du Bellay : leur indice de spécificité proche ou égal à zéro signale que, sur le plan de la simple fréquence d'emploi de *dans*, ils coïncident peu ou prou avec le profil moyen calculé pour ce corpus.

Plusieurs voies d'étude s'ouvrent à l'issue de cette première analyse. On pourrait étudier d'abord quelle attitude d'autres auteurs (Montaigne, par exemple) ont adopté vis-à-vis de l'usage de *dans*. S'inscrivent-ils dans un mouvement plus général qui les amène à modifier leur usage de ces prépositions? Ou bien au contraire reconduisent-ils, texte après texte, des résistances particulières (on songera à Rabelais)? Voire corrigent-ils certaines de leurs premières audaces (on songera à Ronsard)? On aurait tout intérêt à cet égard à comparer les éditions successives de mêmes textes. Une autre direction de travail consisterait à déterminer si, outre les critères temporels et d'auteur, le critère du genre discursif permet aussi d'observer des spécificités remarquables dans l'usage de *dans*. Ainsi pourrait-on observer sous cet angle la place qu'occupe l'œuvre de Ronsard à l'intérieur du genre poétique au XVI^e siècle, ou encore celle de Léry parmi les récits de voyage.

6. Combinatoires préférentielles de *en, dans* et *dedans* : étude quantitative et qualitative

Quels sont les collocatifs¹⁹ préférés (à droite) pour ces trois prépositions au XVI^e s.? Autrement dit, avec quels mots ces prépositions, prises comme pivots, manifestent-elles dans leur usage une « affinité » statistiquement hautement significative? Notre hypothèse est que la combinatoire préférentielle des mots permet de mieux appréhender leurs préférences d'usage, révélant ainsi obliquement certaines de leurs caractéristiques sémantiques profondes. Comme le rappelle Blumenthal (2008 : 37) - reprenant le point de vue de Wittgenstein - « *l'étude de l'usage des mots, réalité manifeste, observable et au besoin quantifiable, est susceptible de nous renseigner sur leur signification – leur nature profonde en quelque sorte* ».

Dans la présente étude, nous restreindrons notre enquête à deux axes :

- i) la combinatoire préférée de *en, dans, dedans* avec les déterminants (§ 6.1.) ;
- ii) les collocatifs nominaux préférés de ces trois prépositions (§ 6.2.)

6.1. *En, dans* et *dedans* : combinaison avec les déterminants

La plateforme TXM permet de construire le tableau des différents cooccurents²⁰ entourant les apparitions au sein du corpus d'un motif exprimé par une requête CQL. Nous avons fait porter le calcul sur le premier mot apparaissant à droite de la préposition prise comme mot-pivot, palliant du même coup (quoique imparfaitement) l'incapacité actuelle de recourir à la catégorie morphosyntaxique

¹⁹ Le terme « collocatif » est entendu dans une acception strictement quantitative, pour désigner tout cooccurent d'un mot pivot qu'un calcul statistique identifie comme hautement spécifique.

²⁰ Sur le corpus Frantext, le calcul de cooccurrences est adapté pour être interprété de la même manière que sur un corpus textuel habituel. On télécharge à partir de la base non catégorisée et pour la période 1500-1599 (tous genres) la liste des fréquences de tous les mots du corpus de travail dont sont tirés les extraits. Cette liste mot-fréquence suffit pour faire le calcul de cooccurrences comme si on disposait des textes entiers.

(absence d'étiquetage du corpus : cf. § 4.) en nous rabattant sur une contrainte d'ordre positionnel et distributionnel. En effet, un tel « fenêtrage » permet de capter les déterminants qui apparaissent dans la subséquence immédiate de la préposition²¹, même s'il demeure possible que le déterminant figure au rang 2 lorsqu'il est précédé d'un prédéterminant. Par exemple :

- (6) *il ne me souvient point d'avoir jamais leu en tous les livres (...)* (G. de Rués, *Les Dialogues de Guy de Brués contre les nouveaux académiciens*, 1557)

Ni *el, ou, on* au singulier ni *es, ès, és* au pluriel, qui amalgament la préposition et le déterminant, ne seront envisagés ici.

Le tableau 4 récapitule les scores de spécificité calculés :

<u><i>dans</i></u>	
Corpus Epistemon (CEp ^{XVI})	Corpus Frantext
Déterminants	Déterminants
[score 127] <i>le</i> <i>les</i> (11), <i>un</i> (11), <i>une</i> (11)	[score 127] <i>le, la, les, un, une, mon</i> <i>son</i> (14), <i>l'</i> (11), <i>leurs</i> (7), <i>ses</i> (6)
<u><i>dedans</i></u>	
Corpus Epistemon (CEp ^{XVI})	Corpus Frantext
Déterminants	Déterminants
[score 127] <i>le</i> <i>la</i> (9), <i>les</i> (8), <i>un</i> (8), <i>l'</i> (6)	[score 127] <i>le, la, l', son, ung</i> <i>mon</i> (16), <i>les</i> (14), <i>ledit</i> (9), <i>une</i> (8), <i>ton</i> (7), <i>ledict</i> (7), <i>ladicte</i> (6), <i>sa</i> (6), <i>un</i> (6), <i>leurs</i> (6)
<u><i>en</i> (Epistemon : CEp^{XVI})</u>	
Déterminants	
[score 127] <i>article défini</i> : <i>la, l'</i> <i>article indéfini</i> : <i>un, une</i> <i>dét. possessif</i> : <i>mon, son, sa, ses, nostre, leur, leurs</i> <i>dét. démonstratif</i> : <i>ce, cest, ceste, ces</i> <i>dét. quantifiant</i> : <i>quelque, plusieurs</i>	

Tableau 4 : récapitulatif des scores de spécificité les plus élevés obtenus pour les cooccurrents placés dans une fenêtre de 1 mot à droite de « *dans* », « *dedans* », « *en* » pris comme mot-pivot (Corpus CEp^{XVI} et Frantext)

Les valeurs numériques doivent être interprétées comme suit : le score de spécificité « 11 » par exemple, affecté au déterminant *les* dans le corpus CEp^{XVI} pour la préposition *dans*, signifie qu'il y a dans ce corpus une probabilité de 10⁻¹¹ que ce mot apparaisse avec la fréquence absolue observée dans la position considérée (c'est-à-dire dans la subséquence immédiate de *dans*). Cette association n'est donc clairement pas fortuite. Ajoutons que la valeur « 127 » proposée dans ce tableau est de pure convention : elle signale un score de spécificité si élevé qu'il outrepassa les capacités de calcul de la machine.

On n'observe pas de différences statistiques significatives concernant les

²¹ La contrainte imposée par le séquençage *en* + *déterminant* exclut le pronom adverbial, normalement suivi d'un verbe dont il est un clitique.

En, dans, dedans prépositions au XVI^e siècle. Approche statistique et combinatoire.

combinatoires de *en, dans* et *dedans* avec les articles définis²², les articles indéfinis *un*²³, *une*, les déterminants possessifs *mon, son*. En revanche, *en* manifeste une affinité très remarquable pour les déterminants démonstratifs *ce, cest, ceste, ces*. Cette affinité a aussi été identifiée par Blumenthal (2012) au XVII^e s., qui y voit le signe d'une forte tendance de *en* à se rattacher à la sphère thématique. Cette tendance se vérifie donc dès le XVI^e s.

Afin d'illustrer cette affinité de la préposition *en* avec la sphère de la thématité²⁴, nous proposons de détailler ci-dessous le fonctionnement en discours (dans notre corpus) de la séquence *en {cet, cest} + {endroit, endroit}* - désormais simplifiée en : *en cet endroit* (*cet* et *endroit* désignant un lemme). Le nom *{endroit, endroit}* constitue en effet le collocatif nominal préféré à droite de la séquence « *en {cet, cest}* » prise comme motif-pivot (calcul de cooccurrence et indice de spécificité). Quant à la séquence envisagée, elle présente l'intérêt de ressortir à la fois à la sphère de la thématité et de pouvoir fonctionner comme un marqueur de topique.

Rappelons d'abord que pour le nom *endroit*, le TLF recense dans le paragraphe final « *étymologie et histoire* » de l'article qu'il lui consacre les sens suivants : 1) a) *lieu déterminé* ; b) *localité où l'on est né, où l'on habite* 2. a) *côté d'une chose qui doit être vu, bon côté* ; b) *égard, point de vue*; 3. a) *partie d'une chose*; b) *passage d'un texte*.

L'étude des corpus CEP^{XVI} et Frantext (XVI^e s.) fait apparaître que :

* le SN *cet endroit* y est presque toujours complément d'une préposition : il s'agit quasi-systématiquement de *en*, jamais de *dans* ou *dedans*.

* Une analyse des valeurs sémantiques de la séquence *en cet endroit* montre une grande rareté du sens de localisation dans l'espace concret. Dans les quelques occurrences observées où la séquence possède cette valeur, l'adverbe *là* se trouve fréquemment postposé au nom *endroit*.

(7) *mais respondit tousjours (monstrant son ventre) qu'il estoit caché en cest endroit là*

* Beaucoup plus souvent, la structure exprime un repérage qui s'exerce sur le *dit* (glose possible : *en cet endroit de mon discours* : cf. (8)), et plus souvent encore sur le *dire* (repérage topical, glose possible : *à ce sujet, à ce propos, sur ce point* ; cf. (9)). Dans le premier cas, l'expression permet au scripteur de localiser *déictiquement* un point ou un passage dans le déroulement de son discours (sens 3.b. du TLF) :

²² Bien entendu, dans le cas de *en*, on ne peut établir de comparaison avec *dans* et *dedans* que pour les articles définis *la* et *l'* puisque pour *le, les*, il se produit un amalgame.

²³ La fréquence relative de la variante graphique *ung* décroît spectaculairement à partir de la décennie 1540-1549. La prééminence de la fréquence relative de *dedans* sur celle de *dans* pour la même période constitue une explication plausible au fait que *ung* apparaît parmi les cooccurents préférés de la première mais non de la seconde.

²⁴ Thématique est à entendre au sens de ce qui, dans un énoncé donné, est « connu » (c'est-à-dire cognitivement et contextuellement actif, ou au moins accessible).

- (8) *Bien escriroys encores aultre chose,
Mais mieulx me vault rendre ma lettre close
En cest endroit : car les Muses entendent
Mon rude stile (...)
(Adolescence Clementine, 1538)*

Dans le second cas, elle permet de rappeler *anaphoriquement* le topique de discours (ce dont il est question dans le passage : sens 2.b. du TLF, et plus précisément : « *relativement à cette question, sur ce point* »). Ainsi, dans (9), *en cest endroit* promeut l'habillement (*des habillemens*) comme topique²⁵ de discours (*au sujet des habillements*).

- (9) *Voila en passant un petit mot de son humanité, et serois contant n'estoit, comme il à esté touché ci dessus, que quand nous eusmes mis pied à terre en son Isle, il nous dit nommément qu'il vouloit que la superfluité des habillemens fut reformee, de finir ici de parler de luy. Il faut doncques que je dise encores le bon exemple et la pratique qu'il monstra **en cest endroit**. Ayant grande quantité tant de draps de laine (qu'il aimoit mieux laisser pourrir dans ses coffres que d'en revestir ses gens, une partie desquels neantmoins estoient presque tous nuds) que de soye (...)*

On observera que dans certains énoncés, il est contextuellement difficile de démêler l'une et l'autre valeurs. Par exemple :

- (10) *Baptiste Leon escrit pareillement que, du temps du Pape Martin cinquiesme, fut trouvé en une grande pierre solide un serpent vif enclos, n'y ayant aucune apparence de vestige par lequel il deust respirer. **En cest endroit** je veux reciter une semblable histoire : estant en une mienne vigne pres le village de Meudon (...)* (A. Paré, *Des monstres et des prodiges*, 1585)

Le locuteur opère-t-il ici un repérage *déictique* ayant pour cible son dit (« à cet endroit de mon propos ») ou bien un repérage *anaphorique* ayant pour cible le topique de son dire (marqueur de topique : « A ce sujet »)? Impossible de trancher.

Dans tous les cas (énoncés (7) à (10)), *en cet endroit* relève bien de la sphère thématique de l'énoncé : qu'il fonctionne déictiquement ou anaphoriquement, l'expression réfère dans le discours à une entité présupposée par le locuteur comme « connue » de son allocutaire.

Signalons pour finir que la séquence « *dans cet endroit* » devient possible au XVII^e s. (quoique marginalement : 27 occ. contre 551 pour *en cet endroit* dans Frantext) et s'accroît quantitativement au XVIII^e s. (148 occ. contre 220 avec *en*). Mais contrairement à *en cest endroit*, elle revêt une valeur majoritairement spatiale,

²⁵ Topique « local », car le topique « global » du discours est la personne de Villegagnon (délocuté dans le texte) dont il est question dans le passage depuis plusieurs lignes déjà.

En, dans, dedans prépositions au XVI^e siècle. Approche statistique et combinatoire.

même s'il lui arrive aussi d'exprimer les autres valeurs illustrées par (7) à (10)²⁶. Nous reviendrons sur ce point dans notre conclusion générale.

6.2. En, dans et dedans : accompagnateurs nominaux préférés à droite

Avec quels noms (situés à droite) les trois prépositions tendent-elles préférentiellement à se combiner ? Pour limiter notre propos, nous nous bornerons aux scores de spécificité affectés aux noms apparaissant à une distance de 1 mot à droite²⁷ après la préposition prise comme mot-pivot.

Le tableau 5 récapitule les résultats obtenus (pour Frantext, seuls les scores maximum - par convention, « 127 » - ont été retenus).

dans	Corpus Epistemon (CEp ^{XVI})	Corpus Frantext (Fr ^{DANS})
	[127] <i>barque, eau, étang, lict, navire</i>	[127] <i>barque, bois, chambre, ciel, {cœur, cœur}, corps, entrailles, eau, lict, logis, maison, mer, pays, poisle, salle, sang, sein, tonneau, vaisseau, veines, ventre, ville</i>
dedans	Corpus Epistemon (CEp ^{XVI})	Corpus Frantext (Fr ^{DEDANS})
	<i>cœur (10), estang (10), corps (10)</i>	[127] <i>bouche, {cœur, cœur}, corps, eau, lict, maison, mer, sein, ventre, ville</i>
en	[Epistemon (CEp ^{XVI})]	
	[127] <i>air, art, cas, chambre, {endroit, endroit, endroits, endroitz}, entendement, esprit, estat, facon, instant, isle, langage, langue, lieu, lieux, main, maison, {maniere, manieres}, monde, partie, pays, place, presence, sorte, temps, terre, village, ville</i>	

Tableau 5 : récapitulatif des noms placés à une distance de 1 mot à droite de « dans », « dedans », « en » pris comme mot-pivot dans les corpus CEp^{XVI} et Frantext (scores de spécificité maximale)

On observe que *dedans* et *dans* possèdent une combinatoire nominale préférentielle extrêmement proche (*dedans* partage 90% de ses collocatifs nominaux avec *dans*), alors que celle de *en* s'en distingue nettement.

Le lexique nominal préférentiellement sélectionné par *dedans* et *dans* possède les caractéristiques suivantes :

- certains noms relèvent du champ lexical de l'eau : *barque, eau, mer, {étang, estang}, vaisseau, navire*

On trouve par ailleurs

- des noms de lieu :

*intérieurs, quotidiens : *chambre, lict, logis, maison, poisle, salle, tonneau ;*

*extérieurs : *bois, ciel, pays, ville ;*

- des noms de partie du corps : *bouche, {cœur, cœur}, corps, entrailles, sang, sein, veines, ventre.*

²⁶ D'où une étude plus fine qu'il serait utile de conduire pour déterminer quelles différences on observe dans la distribution de ces deux expressions au XVII^e s. notamment, dans leurs emplois spatiaux aussi bien que non spatiaux.

²⁷ c'est-à-dire séparés de la préposition par un mot (presque toujours un déterminant). Nous avons par conséquent écarté les emplois de *en* à régime nominal sans déterminant, qu'il conviendra d'étudier (et de comparer avec les emplois à régime nominal déterminé) dans un autre travail.

Pour la préposition *en*, on observe certes aussi des noms de lieu :

- intérieurs, quotidiens : *chambre, maison* ;
- extérieurs : *isle, monde, pays, place, terre, village, ville* ;

ou de partie du corps : *main*.

Mais ce qui retient surtout l'attention, c'est la forte présence de termes généralement abstraits ou génériques dénotant (sens premier)

- un lieu : {*endroit, endroict, endroits, endroictz*}, {*lieu, lieux*}, *place*
- un contenant : *air*
- un domaine : *art*
- un intervalle temporel : *temps, instant*
- une entité relative à l'esprit humain: *entendement, esprit, langage, langue*
- une notion très générale : *cas, estat, facon, {maniere, manieres}, partie, presence, sorte*.

Ces résultats, à considérer avec une certaine prudence du fait des limites actuelles de notre corpus, soulèvent d'ores et déjà des questions d'un grand intérêt.

Concernant *dans* et *dedans* : les préférences combinatoires identifiées varient-elle suivant le genre discursif ? Le cas des noms du corps en particulier est suggestif : la poésie y recourt-elle plus que les autres genres ?

Concernant *en*, nous ne ferons que quelques observations en guise de remarque finale. D'abord, la préférence marquée de cette préposition pour les noms génériques et abstraits apparaît significative dès le XVI^e siècle et ne semble donc pas (du moins directement) liée à l'accroissement spectaculaire des emplois de *en* avec régime nominal sans déterminant qu'on observe à partir du siècle suivant et qui ira croissant jusqu'à nos jours (Blumenthal, 2011 : 295). En second lieu, on observe que dès cette période, l'expression figée *en l'air* est déjà très courante alors que *dans l'air* et *dedans l'air* sont rares. Enfin, il est très remarquable que seul le nom de partie du corps *main* se trouve en combinaison préférentielle avec *en*, tandis qu'on ne trouve pas d'affinité statistiquement significative de ce nom avec *dans* ou de *dedans* – à l'inverse des noms {*cœur, cueur*}, *corps sang, sein, ventre*. Si l'on y regarde de près, on constate que la fréquence absolue de la séquence *en* + (*det* + *E*²⁸) + *main* l'emporte ... haut la main dans CEP^{XVI} comme dans CF_r^{DANS} et CF_r^{DEDANS}, avec beaucoup d'emplois en complément des verbes *tenir, porter, mettre, prendre, avoir*. On avancera ici l'hypothèse que dès le XVI^e siècle, *en* inclut dans son identité sémantique un trait « *praxique* » qui la fait déjà préférer à *dans* (qui reste cependant toujours possible) lorsqu'on veut exprimer un contrôle (Leeman (1995 : 57), Cadiot (1997 : 198)) opéré par un agent humain sur l'entité régime en vue d'une action. Ainsi, si l'on (*tient* + *porte* + *met* + *a* + ...) quelque chose *en* + (*det* + *E*) + *main*, c'est avant tout pour agir en contrôlant cet objet :

(11) *porte a son col en echarpe une trousse, et en sa main senestre ung arc bende* (G. Tory, *Champ fleury*, 1529)

(12) *Il te faut maintenant en main les armes prendre* (P. de Ronsard, *Remonstrance au peuple de France*, 1563)

²⁸ « E » représente l'élément neutre de la concaténation et sert à marquer la séquence vide (i.e. l'absence de déterminant, ici). (Molinier & Levrier, 2000 : 22).

En, dans, dedans prépositions au XVI^e siècle. Approche statistique et combinatoire.

Alors qu'avec (*dans* + *dedans*) + *det.* + *main* – séquence beaucoup plus rare – l'idée de contrôle ne s'impose pas. D'où des énoncés où la séquence permet la simple localisation d'une cible sans contrôle opéré (par la main) sur elle en vue d'une action. Par exemple²⁹ :

(13) *Le Temps la jecta dans la main*

Du filz de Saturne, et de Rhée

(P. de Ronsard, *Ode à Michel de l'Hospital*, 1550)

(14) *devant eulx mist sa lettre dedans la main du mort* (Anonyme, *Le Violier des histoires rommaines moralisées*, 1521)

Un siècle plus tard – faut-il s'en étonner ? – Malherbe fera cette observation : « *On ne dit pas : il avait l'arc (...) dans la main mais en la main* ». (F. Brunot cité par A. Sancier-Château, *ibid* : 226)

7. Conclusion

Nous avons montré que, dans les textes littéraires, les courbes d'évolution des fréquences relatives de *dans* et *dedans* au XVI^e siècle suivent deux trajectoires opposées, la première allant croissant avec une accélération spectaculaire après 1560, tandis que la seconde accuse un fléchissement global avec une chute particulièrement nette dans la première moitié du siècle. Concernant *dans*, notre approche statistique a permis de confirmer l'analyse selon laquelle, dans la langue littéraire, avec Ronsard tout change. Le calcul de spécificité appliqué à la partition « Auteurs » nous a en outre permis d'aller plus loin, en situant d'une part l'ensemble des auteurs du corpus Epistemon les uns par rapport aux autres (relativement à l'emploi de *dans*), en révélant d'autre part la position originale occupée par Léry (*Histoire d'un voyage fait en la terre du Bresil*, 1578). La forte fréquence observée de *dans* dans ce texte est-elle le fruit d'une innovation ? Ou bien ne manifeste-t-elle *in fine* qu'une récurrence significative d'emplois par ailleurs déjà banals depuis Ronsard voire avant, mais qui apparaissent statistiquement significatifs du fait du genre discursif lui-même (la relation de voyage) ? Questions qui ouvrent vers d'autres interrogations. Quelle part d'innovation trouve-t-on dans l'œuvre de Ronsard par rapport à ses devanciers ? Innovation dans la combinatoire morphosyntaxique de *dans* ? dans sa combinatoire lexicale ? Ronsard avait-il eu des prédécesseurs ? Un élargissement de notre corpus serait ici extrêmement précieux. L'inclusion d'autres textes relevant d'autres genres discursifs modifierait à coup sûr le paysage des spécificités. Rappelons enfin que nous avons identifié chez Rabelais un hapax qui, jusqu'à aujourd'hui, n'avait pas retenu l'attention et qui, comme nous avons tenté de le montrer, revêt un certain intérêt.

Comme il a été signalé, l'exploration statistique des combinatoires de *dans*, *dedans* et *en* est actuellement entravée par le fait que nous ne disposons pas, pour les textes de notre corpus, d'un étiquetage morphosyntaxique ni d'une lemmatisation satisfaisants. Nonobstant, plusieurs éléments dignes d'intérêt apparaissent d'ores et déjà. D'abord, *en*, *dans* et *dedans* ne se distinguent pas significativement entre eux relativement à l'emploi des articles définis (*la*, *l'* pour *en*), des articles indéfinis singuliers *un*, *une* et de certains possessifs. On observe en revanche une affinité forte

²⁹ Dans (13), *dans la main* explicite simplement le lieu d'arrivée de la cible jetée. Idem dans (14).

de *en* – lorsque la préposition est suivie d'un régime nominal déterminé - avec les déterminants démonstratifs. Comme le souligne Blumenthal (2012), cela plaide pour une affinité de cette préposition avec le pôle de la thémativité. Enfin, il apparaît clairement que *en*, contrairement à *dans* et *dedans*, manifeste une préférence combinatoire pour certains termes généraux et abstraits. Ces deux dernières observations, combinées aux résultats de l'analyse que nous avons menée pour l'expression *en cet endroit*, suggèrent une hypothèse. Au XVI^e siècle, *en* se distinguerait de *dans* et *dedans* par sa capacité à entrer dans la composition d'expressions à sens nettement plus abstrait, parmi lesquelles des expressions métadiscursives à portée sur le dire ou le dit. A l'inverse, *dans* et *dedans* participeraient le plus souvent à la construction du *sens référentiel* des énoncés. Si cette hypothèse était vérifiée par l'étude exhaustive des expressions statistiquement significatives dans lesquelles entrent *en*, *dans* et *dedans* au XVI^e siècle, il serait alors intéressant de se demander si la perte progressive de la capacité de *en* à être suivie d'un régime avec déterminant n'aurait pas contribué à accélérer à partir du XVII^e siècle ce mécanisme de spécialisation de *en* déjà lancé à la Renaissance.

Denis Vigier (Université de Lyon2
UMR ICAR (CNRS, Lyon2, ENS, IFÉ)

Bibliographie

- Blumenthal, P. (2008). « Combinatoire des prépositions: approche quantitative », *Langue Française*, 157, 37-51. Paris: Larousse.
- Blumenthal, P. (2011). Le figement: du XVII^e siècle au français contemporain. Dans: J.-C. Anscombre et S. Mejri (dir.), *Le figement linguistique: la parole entravée*. Paris: Éditions Champion, pp 283-302.
- Blumenthal, P. (2012, à paraître). Évolutions de la combinatoire prépositionnelle: le cas de *en*, Dans: *Mélanges à Olivier Soutet*.
- Brunot, F. (1906). *Histoire de la langue française. Des origines jusqu'en 1900*. Tome II: *le seizième siècle*. Paris: A. Colin.
- Cadiot, P. (1997). *Les prépositions abstraites en français*. Paris: A. Colin.
- Darmesteter, A. (1885). *Notes sur l'histoire des prépositions françaises en, enz, dedans, dans*. Paris: Le Cerf.
- De Mulder, W. (2008). « *En* et *dans*: une question de « déplacement » ? ». Dans: O. Bonami, S. Prévost, M. Charolles, J. François et C. Schnedecker (dir.), *Discours, diachronie, stylistique du français: études en hommage à Bernard Combettes*. Bern: Peter Lang, 277-291.
- Fagard, B., Sarda, L. (2009). « Etude diachronique de la préposition *dans* ». Dans: *Autour de la préposition: position, valeurs, statut et catégories apparentées à travers les langues*, J. François, E. Gilbert, C. Guimier et M. Krause (dir.), *Bibliothèque de Syntaxe et Sémantique*. Caen: PUC, pp 225-236.
- Frautschi, R. L. (1963). The 'énigme en prophétie' (« *Gargantua* LVIII) and the question of authorship, *French Studies*, XVII, 4, 331-340. Oxford: Oxford University Press.

En, dans, dedans prépositions au XVI^e siècle. Approche statistique et combinatoire.

- Gougenheim, G. (1950, 1970). « Valeur fonctionnelle et valeur intrinsèque de la préposition « en » en français moderne ». Dans: *Etudes de grammaire et de vocabulaire français*, réunies sur l'initiative de ses collègues et amis pour son soixante-dixième anniversaire. Paris: Picard, pp. 55-65.
- Gougenheim, G. (1945). « Les prépositions « en » et « dans » dans les premières œuvres de Ronsard ». Dans: *Etudes de grammaire et de vocabulaire français*, réunies sur l'initiative de ses collègues et amis pour son soixante-dixième anniversaire. Paris: Picard, pp. 55-65.
- Gougenheim, G. (1951, 1974). *Grammaire de la langue française du seizième siècle*. Paris: Picard
- Guillot, C., Heiden S., Lavrentiev A. (2007). « Typologie des textes et des phénomènes linguistiques pour l'analyse du changement linguistique avec la Base de Français Médiéval ». Dans: D. Malrieu (dir.), *Corpora et Questionnements du littéraire, Actes des Journées d'étude internationales*. LINX, numéro spécial, 125-139. Paris: Centre de recherches linguistiques.
- Haase, A. (1969). *Syntaxe française du XVII^e siècle*, traduite et remaniée par M. Obert. Paris: Delagrave.
- Heiden, S. (2010). « The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. » Dans: *24th Pacific Asia Conference on Language, Information and Computation*, I. Kiyoshi, R. Otaguro (dir). Institute for Digital Enhancement of Cognitive Development, Waseda University. pp. 389-398.
- Huchon, M. (1981). *Rabelais grammairien. De l'histoire du texte aux problèmes d'authenticité*. Genève: Droz.
- Leeman, D. (1995). « Pourquoi peut-on dire *Max est en colère* mais non **Max est en peur* ? Hypothèses sur la construction *être en* ». *Langue française*, 105, 55- 69. Paris: Larousse.
- Pincemin B. (1999), *Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*, Thèse de Doctorat, Université Paris IV (Sorbonne).
- Sancier-Château, A. (1995). *Une esthétique nouvelle: Honoré d'Urfé correcteur de L'Astrée, (1607-1625)*. Genève: Droz
- Spang-Hanssen, E. (1963). *Les prépositions incolores du français moderne*. Copenhague, G.E.C GAD Forlag.
- Vigier, D. (2013), « Sémantique de la préposition *en* : quelques repères », *Langue Française*, 178. Paris : Larousse.

Résumé

Cet article, réalisé à partir d'analyses statistiques, étudie la concurrence entre les prépositions *en, dans* et *dedans* au XVI^e siècle. Après avoir étudié l'évolution quantitative des emplois de *dans* et *dedans* au cours de cette période, on examine si *en, dans* et *dedans* peuvent être distingués à partir des combinaisons statistiquement préférentielles dans lesquelles elles entrent. Les corpus utilisés sont d'une part un ensemble de textes littéraires du XVI^e siècle tirés de la « Bibliothèque Virtuelle Humaniste », d'autre part un ensemble d'extraits tirés de la base Frantext (période

1500-1599, tous genres). L'outil automatique mobilisé est la plateforme TXM développée à l'ENS Lyon.

Mots clefs: Préposition "*en*"; Préposition "*dans*"; Diachronie; Combinatoire; Linguistique quantitative.

Abstract

This article, based on statistical analyses, examines the competition between the French prepositions “en”, “dans” and “dedans” in the sixteenth century. It first seeks to study the quantitative evolution of the uses of “dans” and “dedans” during that century. It goes on to examine whether the statistically preferred combinatories of “en”, “dans” and “dedans” can distinguish between these three prepositions. This study is based on two corpora : a set of literary texts from the "Bibliothèque Virtuelle Humaniste" (Corpus Epistemon) and the Frantext database (1500 to 1599, all genres). The automatic tool used is the platform TXM (Textométrie) developed at the ENS - Lyon.

Keywords: French preposition "*en*"; French preposition "*dans*"; Diachrony; Word combination; Quantitative linguistics.