



HAL
open science

La préposition dans au XVI^e siècle. Apports d'une linguistique instrumentée

Denis Vigier

► **To cite this version:**

Denis Vigier. La préposition dans au XVI^e siècle. Apports d'une linguistique instrumentée. *Langages*, 2017, Du quantitatif au qualitatif en diachronie : prépositions françaises, 206 (2), pp.105-122. 10.3917/lang.206.0105 . halshs-01581181

HAL Id: halshs-01581181

<https://shs.hal.science/halshs-01581181>

Submitted on 21 Apr 2020

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.

La préposition *dans* au XVI^e siècle. Apports d'une linguistique instrumentée

The preposition "*dans*" in the 16th century. Contribution of an equipped linguistic analysis

Denis Vigier

Université de Lyon / UMR ICAR (CNRS, Université Lyon 2, ENS de Lyon)

Résumé

La préposition *dans* est récente en français. Son usage s'est considérablement développé à partir de 1550 environ, du moins dans les textes littéraires, pour devenir très vite une des prépositions les plus fréquentes du français. Pourquoi un tel succès ? Dans cet article, nous proposons d'évaluer l'hypothèse formulée jadis par A. Darmesteter (1885) pour expliquer cette subite fortune de *dans*. Notre démarche s'appuie sur une étude statistique des préférences combinatoires que cette préposition manifeste au sein du corpus de diachronie du français Presto.

Mots-clés : préposition dans – diachronie – français préclassique – statistique – textométrie

Abstract

The preposition *dans* is recent in French. Its use increased considerably since about 1550, at least in literary texts, and it quickly became one of the most common French prepositions. Why such a success? In this article, we evaluate the hypothesis formulated by A. Darmesteter (1885) to explain this sudden fortune of *dans*. Our approach is based on a quantitative and qualitative analysis of data collected in the diachronic French corpus Presto.

Keywords: French prepositions *dans* – diachrony – pre-classical French – statistical analysis – textometry

1. INTRODUCTION

On peut considérer pour acquis (voir les nombreuses études sur corpus qui toutes s'accordent sur ce point : Darmesteter 1885 ; Gougenheim 1945, 1951 ; Brunot 1967 ; Fagard & Sarda 2009 ; etc.) que c'est à partir du début de la seconde moitié du XVI^e s. que *dans* fait une entrée remarquable sur la scène des prépositions du français – du moins dans le discours littéraire. C'est en effet autour de 1550 que l'usage de cette préposition – qui auparavant « traîn[ait] une existence obscure » (Darmesteter, 1885 : 185) – s'accroît de manière spectaculaire, *dans* se voyant ainsi appeler « au plus brillant succès » (Darmesteter, 1885 : 185) pour figurer finalement, en français contemporain, au cinquième rang en termes de fréquence d'emploi derrière *de*, *à*, *en* et *pour*¹. Comment expliquer une fortune si étonnante et si rapide ?

Le présent article vise à réexaminer la thèse qu'avait formulée jadis A. Darmesteter pour expliquer l'impressionnante montée en puissance fréquentielle de *dans* à partir de 1550 environ. Selon lui, au cours de la seconde moitié du XVI^e s, « *dans* est venu prendre la place laissée vide par la disparition de *ou* et de *es*, contractions de *en* et de *le* et *les* » (1885 :183). Dans une première partie, nous brosserons un tableau d'ensemble de l'évolution quantitative des usages de *dans* et de *en* entre 1551 et 1940. Puis nous évaluerons l'hypothèse d'A. Darmesteter proprement dite en nous appuyant sur une étude statistique des préférences combinatoires² que cette préposition manifeste en français préclassique et au-delà. Nous serons ainsi conduits à rejeter une partie de cette hypothèse et à proposer une autre voie d'explication possible à la subite fortune de *dans*.

2. TABLEAU GENERAL DE L'EVOLUTION QUANTITATIVE DES USAGES DE *EN* ET DE *DANS* ENTRE 1551 ET 1940

Avant de nous pencher sur la période temporellement limitée où l'on assiste au premier accroissement significatif de la fréquence d'usage de *dans* en français dans le discours littéraire, qu'on nous

¹ Voir liste des 1063 mots les plus fréquents du français oral élaborée par le Centre du Français Élémentaire (Gougenheim *et al.* 1956) et Vaguer (2008 : 23).

² Je tiens à remercier B. Pincemin pour l'ensemble de ses conseils et de ses remarques relatifs aux calculs statistiques mobilisés dans cet article.

permette de brosse dans un tableau général l'évolution des usages de *en* et de *dans* sur cinq siècles environ. Pour ce faire, nous recourons au calcul des spécificités³ disponible sur la plateforme open source TXM⁴ (calcul mis au point par P. Lafon (1980, 1984)) en l'appliquant à une partition⁵ opérée sur⁶ Presto¹⁵⁵¹⁻¹⁹⁴⁰. Précisons cursivement en quoi consiste ce calcul.

2.1. Le calcul des spécificités

Le calcul des spécificités implémenté dans TXM permet de porter un jugement statistique sur la sous-fréquence f_{ij} d'une forme i dans une partie j d'un corpus, étant connus par ailleurs les paramètres suivants :

- F_i : fréquence de la forme i dans le corpus ;
- t_j : taille de la partie j ;
- T : taille du corpus.

Les valeurs f_{ij} , F_i , t_j , T ayant été définies, est alors calculé un indice de spécificité dont la valeur numérique peut être de signe positif ou négatif. Plus la valeur absolue de cet indice est élevée, plus la chance que la sous-fréquence f_{ij} observée pour la forme i dans la partie j soit due au hasard est faible⁷. Un indice supérieur ou égal à 3 signale que la fréquence f_{ij} observée est significativement élevée par rapport à ce que le modèle hypergéométrique laissait prévoir. On dira donc que la forme i est *spécifique positive* pour la partie j . Inversement, un indice inférieur ou égal à -3 signale que la forme est sous-représentée par rapport à ce que le modèle laissait prévoir : on parlera de *spécificité négative*. Enfin, un indice situé entre -3 et +3 sera considéré comme

³ Pour une présentation détaillée du calcul des spécificités, voir Lebart et Salem (1994 : 172-185).

⁴ <http://textometrie.ens-lyon.fr>.

⁵ Par *partition* d'un corpus de textes, il faut entendre sa « division (...) en parties constituées par des fragments de texte consécutifs, n'ayant pas d'intersection commune et dont la réunion est égale au corpus. » (Lebart & Salem, *ibid* : 316)

⁶ Par convention, les dates figurant en exposants après le nom Presto stipulent les bornes temporelles du sous-corpus découpé dans le corpus intégral.

⁷ Le calcul des spécificités de Lafon est fondé sur la loi hypergéométrique, loi de probabilité correspondant à un tirage sans remise dans l'hypothèse d'indépendance (voir Lebart & Salem, 1994 : 174 ; Labbé & Labbé : 1994).

inférieur au seuil de significativité statistique⁸ et la forme *i* sera dite *banale* pour la partie *j*.

Pour le calcul visé, nous avons partitionné le corpus Presto^{1551_1940} en treize parties d'empan temporel identique, soit trente années. Notre objectif consistait à déterminer si les prépositions *en* et *dans* s'avéraient, dans l'une ou l'autre de ces parties, spécifiques positives, spécifiques négatives ou banales. Voici les paramètres définis pour ce calcul :

- f_{ij} = sous-fréquence de la préposition sélectionnée pour le calcul (*en* ou *dans*) dans chacune des treize parties;
- F_i = fréquence de cette même préposition dans le corpus total (Presto^{1551_1940}) ;
- t_j = sous-fréquence de tous les mots⁹ dans chacune des treize parties ;
- T = fréquence tous les mots dans Presto^{1551_1940}.

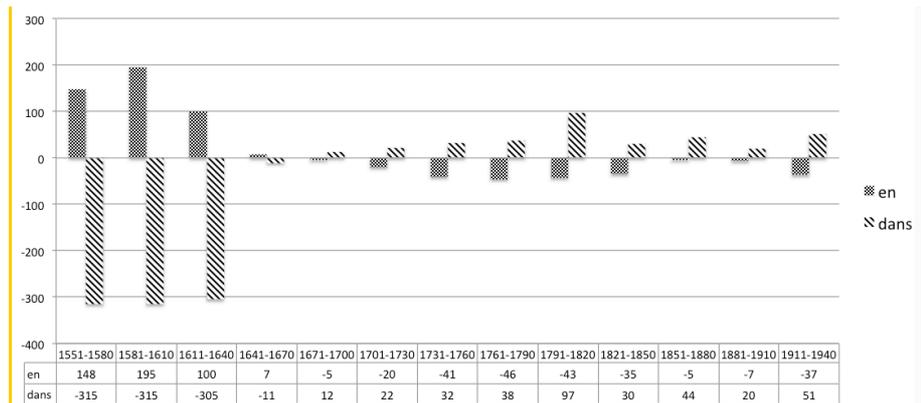
2.2. Présentation et analyse des résultats

Voici les résultats obtenus.¹⁰ Au diagramme est associée la table des valeurs (scores de spécificités) calculées.

⁸ Pour une discussion, voir Reutenauer (2002 : 161).

⁹ Par « mot », il faut entendre toute unité résultant de l'opération de segmentation automatique du flux textuel accomplie sur le corpus Presto.

¹⁰ A. Salem (1988) a montré que le calcul des spécificités appliqué à des suites textuelles chronologiques pouvait gagner à être complété par le recours à des indicateurs (triangle des spécificités connexes, coefficient de Von Neumann, ...) permettant de mieux appréhender la ventilation d'une forme dans les *n* parties du corpus. Ces indicateurs s'imposent lorsque le diagnostic de spécificité porté sur la sous-fréquence d'un terme dans une ou plusieurs parties d'une série chronologique aboutit à un jugement de banalité. Notre étude n'entre pas dans ce cas de figure puisque aucune sous-fréquence de *en* ou de *dans* n'apparaît comme statistiquement banale dans notre partition.



Évolution des scores de spécificité affectés aux prépositions *en* et *dans* entre 1551 et 1940. Corpus Presto, partitionné en 13 tranches de 30 ans

L'examen des spécificités positives et négatives dans ce tableau conduit aux interprétations suivantes : tandis que *dans* voit ses emplois gagner un terrain considérable en cinq siècles¹¹, ceux de *en* s'érodent tout aussi considérablement. *Dans* connaît en outre un accroissement fréquentiel très net à la fin de la première moitié du XVII^e s., comme le montre la chute de la valeur absolue de son score de sous-spécificité pour la période 1641-1670. Parallèlement, *en* connaît une baisse continue de sa fréquence d'emploi, avec une chute particulièrement sensible entre 1611 et 1670. Enfin, la période 1641-1700 apparaît comme charnière puisque *dans* et *en* voient leurs zones de spécificités s'inverser. À l'aube du XVIII^e s., le sort des deux

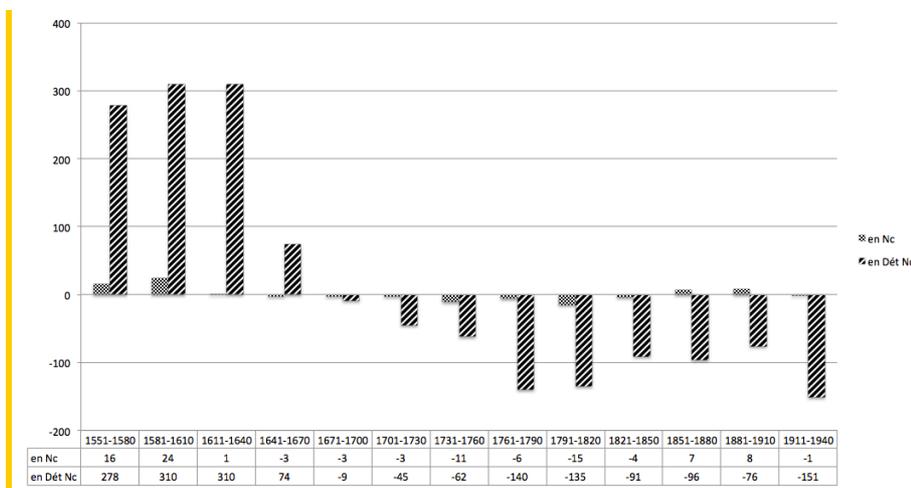
¹¹ Cette tendance générale ne doit pas masquer des variations plus locales des valeurs du score calculé, pour lesquelles des études de détail seraient nécessaires et qui mettraient au jour certains effets de corpus. Par ex., le pic de +97 pour la période 1791-1820 s'explique en partie par la présence, dans cette tranche chronologique, d'un ouvrage de médecine (R. T. Laennec, *De l'auscultation médiate*) qui traite de l'invention du stéthoscope. Comme chaque bruit entendu *via* ce nouvel instrument doit être systématiquement localisé (concrètement – *dans* un organe – ou abstraitement – *dans* une maladie), le scripteur est conduit à sur-utiliser la préposition *dans* par rapport aux autres textes figurant dans cette même période, et même dans une période plus large (1750-1850). D'où une augmentation significative des emplois de *dans* pour cette tranche qui se traduit par un accroissement local du score de sur-spécificité.

prépositions est scellé : *dans* passe définitivement dans la zone de spécificité positive et *en* dans la zone négative.

On peut examiner de plus près la trajectoire des emplois de *en* dans notre corpus en distinguant les cas où il est suivi d'un déterminant et d'un nom commun (*En Det Nc*) versus d'un nom commun nu (*En Nc*). Pour ce calcul, les paramètres sont :

- f_{ij} = sous-fréquence observée dans chacune des treize parties de la séquence S_1 « en suivi d'un déterminant et d'un nom commun » ou de la séquence S_2 « en suivi d'un nom nu »;
- F_i = fréquence de la séquence S_1 ou de la séquence S_2 dans le corpus total (Presto¹⁵⁵¹⁻¹⁹⁴⁰) ;
- t_j = sous-fréquence de tous les mots dans chacune des treize parties ;
- T = fréquence de tous les mots dans Presto¹⁵⁵¹⁻¹⁹⁴⁰.

Voici le diagramme des spécificités auquel on aboutit :



Évolution des scores de spécificité affectés aux séquences *En Det Nc* et *En Nc* entre 1551 et 1940. Corpus Presto, partitionné en 13 tranches de 30 ans

Les évolutions des séquences *En Det Nc* et *En Nc* sont clairement distinctes. La première se voit affecter des scores de spécificités positifs très élevés entre 1551 et 1670 (sur-représentation prononcée), puis des scores significativement négatifs à l'aube du XVIII^e s. Les scores de la seconde séquence oscillent quant à eux autour de l'axe

des abscisses, leur valeur étant alternativement positive puis négative et demeurant à partir du début du XVII^e s. dans une frange de valeurs absolues qui excède rarement 10. En d'autres termes, tandis que la séquence S₁ « en suivi d'un déterminant et d'un nom commun » subit une chute très sévère dans son usage au cours de la seconde moitié du XVII^e s., l'emploi de *en* suivi d'un nom commun non déterminé, quoique soumis lui aussi à une certaine désaffection entre 1640 et 1850 environ, se maintient dans une amplitude de variation fréquentielle limitée et voit même son score de spécificité se redresser au cours du XIX^e siècle.

Récapitulons : l'étude quantitative du corpus Presto¹⁵⁵¹⁻¹⁹⁴⁰ montre que *dans* y connaît un accroissement majeur de ses emplois vers la fin de la première moitié du XVII^e s. A l'inverse, la préposition *en* dont l'emploi s'érode globalement entre le début du XVII^e s. et nos jours, voit son usage avec un régime nominal déterminé subir un ralentissement spectaculaire au cours du XVII^e s. Suivie d'un nom commun *nu*, sa fréquence d'emploi connaît en revanche un ralentissement nettement moindre avec même une phase d'embellie au cours du XIX^e s. qu'il faudrait confirmer sur d'autres corpus.

On sait enfin grâce à de nombreuses études sur corpus concordantes (cf. notre introduction), que l'usage de *dans* a commencé à s'accroître de manière remarquable dès 1550¹², du moins dans le discours littéraire. C'est à cette première hausse fréquentielle que nous allons maintenant nous intéresser.

3. FORTUNE DE *DANS* A PARTIR DE 1550 : L'HYPOTHESE DE DARMESTETER (1885)

L'hypothèse avancée par A. Darmesteter (1885) pour expliquer le surprenant accroissement fréquentiel des usages de *dans* au sein des textes littéraires à partir de 1550 environ s'enracine dans une étude des amalgames issus de la combinaison de *en* avec les formes *le*, *les* de l'article défini au XVI^e s.

¹² Ce premier « saut » demeurant insuffisant en termes de fréquences relatives pour être sensible dans les calculs de spécificités sur les partitions dont les résultats ont été présentés plus haut (diagramme 1).

3.1. Bref rappel de la situation des amalgames issus de la combinaison de *en* avec les formes de l'article défini *le* et *les* au XVI^e siècle

Dès le Moyen Âge, *en* placé devant les formes de l'article défini *le* et *les* se contracte pour donner des formes amalgamées. En revanche, il ne se produit aucune contraction devant les formes *la* et *l'* de ce même article ni devant les autres déterminants. Une telle situation est parallèle à celle observée pour les prépositions *de* et *à* combinées avec les mêmes formes de l'article défini.

Les formes amalgamées morphologiquement issues de *en* + *le/les* qu'on observe dans les textes entre les IX^e et XVI^e s. sont nombreuses. En partant des dictionnaires de F. Godefroy (1891-1902) et E. Huguet (1925-1967) et de la grammaire de S. Lardon & M.- C. Thomine (2009), on peut proposer la liste suivante :

- *en le > enl, el, eu, u, o, hou, hu, ou, on, om, un
- *en les > ens, ans, eins, ons, eis, eus, aus, as, es, ès, és, ez

A l'aube du XVI^e s. se produit un fait majeur : la disparition progressive des amalgames morphologiquement issus de *en* (entamée bien avant 1500) entraîne leur remplacement « soit par la préposition *dedans*, soit par *a* contracté avec l'article masculin : *au* (qui sonne presque comme *ou* = o) et *aux*. » (Brunot, 1967 : 278). La situation devient la suivante : « *Au* et *aux* se trouvaient donc correspondre à la fois à *à* + *le*, *à* + *les* d'une part et à *en* + *le*, *en* + *les* d'autre part. » (Gougenheim, 1951-1974 : 182).

Voici quelques exemples illustrant non seulement l'ambiguïté potentielle des morphèmes *au*, *aux* mais aussi certains « flottements » (Brunot 1967) entre *es* et *aux* notamment chez Rabelais :

- (1) Face le ciel (quand il voudra) revivre / Lisippe, Apelle, Homere, qui le pris / Ont emporté sur tous humains esprits / **En la** statue, **au** [= *en le] tableau, et **au** [= *en le] livre. (1550, Du Bellay, *L'Olive*) (cité par Lardon & Thomine, 2009 : 393)
- (2) **Aux** [= *en les] braves exploits de sa vie, et **en sa** mort, on le [=Caton] sent toujours monté sur ses grands chevaux. (Montaigne, Essais, III) (cité par Gougenheim, 1951-1974 : 183)
- (3) De tant loing que le vit Pantagruel, il dist **es** [= *à les] assistans : Voyez vous cest homme qui vient

par le chemin du pont Charanton ? (1542, F. Rabelais, *Pantagruel*)

Une telle situation de « surcharge » (Gougenheim 1951-1974) aurait créé diverses confusions susceptibles d'entraver l'interprétation de certains énoncés. Ainsi dans :

(4) Amour ne change point le cueur, mais le monstre tel qu'il est, fol **aux** folles et saige **aux** saiges. (1550, M. de Navarre, *l'Heptameron* ; cité par Gougenheim, 1951-1974 : 183)

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* au sens moderne de *chez* : *fou chez les fous, sage chez les sages* ?

3.2.L'hypothèse d'A. Darmesteter

Cette thèse formulée en 1885 a été reprise par G. Gougenheim (1951), F. Brunot (1967) et L. Terreaux (1968) entre autres, et prévaut encore dans la littérature linguistique actuelle¹³ sans être cependant toujours formulée dans ses détails. On peut y distinguer deux volets. Le premier a trait à la *cause* de la fortune fréquentielle de *dans* à la fin de la première moitié du XVI^e siècle : d'après l'auteur, la préposition *dans* aurait pris la place laissée vide par la disparition des amalgames morphologiquement issus de **en le*, **en les* provisoirement remplacés par les morphèmes *au*, *aux*.

« Là où nous employons *dans*, le Moyen Âge disait *ou*, *es* : *ou champ*, *es champs*, *es circonstances*. Ainsi *en* s'est maintenu dans les cas où il n'y avait pas lieu de le combiner avec l'article ; *dans* s'est substitué à *en* dans ceux où *en* se contractait avec *le* et *les*. (...) *Dans* est venu prendre la place laissée vide par la disparition de *ou* et de *es*, contractions de *en* et de *le* et *les*. (...) Il y a coïncidence entre la disparition de *ou* et *es* et le développement extraordinaire acquis par *dans*. L'une est la cause de l'autre, il n'est pas difficile de le prouver » (A. Darmesteter, *ibid.* : 184).

¹³ Voir par ex. W. de Mulder (2008 : 287) : *La plupart des auteurs consultés établissent un rapport entre la création de la préposition dans et l'évolution des formes de en devant l'article défini.*

Cette preuve (c'est le second volet), l'auteur la tire de l'étude quantitative de la distribution des déterminants placés dans le régime de *dans* au sein de quelques textes de la seconde moitié du XVI^e s. En effet, déclare A. Darmesteter, « [Si] *dans* s'est substitué à *en* dans [les emplois] où *en* se contractait avec *le* et *les*, [alors] les premiers emplois ont dû être ceux où *dans* était suivi de l'article *le* et d'un mot commençant par une consonne, ou de l'article *les* » (*ibid*). Ainsi accomplit-il un relevé des occurrences de cette préposition dans le premier volume de l'édition Blanchemain des *Amours* de P. de Ronsard.: « Les exemples [tirés] de Ronsard confirment cette vue [= notre hypothèse], puisque sur cinquante-quatre exemples, *dans* est suivi de *le* ou *les* dans trente cas et d'un autre mot quelconque dans vingt-quatre cas seulement¹⁴ » (185, 186).

Soucieux d'étayer davantage cette thèse, les successeurs d'A. Darmesteter se sont employés à approfondir la piste quantitative en l'enrichissant de nouveaux relevés accomplis manuellement sur corpus. Ainsi G. Gougenheim (1950) présente-t-il les résultats d'un dépouillement conduit sur les quatre premiers volumes des œuvres de Ronsard (collection de la *Société des Textes français modernes*) qui le conduisent à conclure que cet auteur serait le premier à opter préférentiellement pour *dans* plutôt que pour *en* devant *le* et *les*. L. Terreaux élargit le champ du dépouillement à toute l'œuvre de Ronsard et conclut dans le même sens (voir en part. 1968 : 195-204).

Il reste que ces enquêtes, pour minutieuses qu'elles fussent, sont demeurées limitées dans leur ampleur : jusqu'à l'arrivée de l'outil informatique, le dépouillement manuel s'avérait une tâche extrêmement coûteuse en temps, donc limitée en termes de nombre de mots traités et non exempte d'erreur¹⁵. Avec l'avènement de l'informatique et la miniaturisation de ses composants, l'étude systématique du vocabulaire menée sur des corpus de plus en plus vastes peut s'effectuer aujourd'hui en quelques millisecondes et se trouve à la portée de tous les chercheurs. Il est ainsi possible de tester sur un corpus de grande ampleur l'hypothèse d'A. Darmesteter.

¹⁴ On voit ici se profiler la faiblesse du raisonnement : l'auteur se fonde sur des fréquences absolues sans poser la question de la distribution des articles définis dans l'ensemble du corpus, quelle que soit leur position. Il va de soi qu'à une époque où les chercheurs ne disposaient pas de l'outil informatique, une telle question ne pouvait demeurer que sans réponse.

¹⁵ Voir D. Vigier (2015).

Une telle étude n'est en rien mineure ou anecdotique. Il s'agit de mieux comprendre par quelles voies une préposition du français contemporain a pu s'imposer récemment et de manière fulgurante dans le peloton de tête de ses mots les plus fréquents. Préposition faisant par ailleurs figure d'exception parmi les langues romanes puisque seul, parmi elles, « le français moderne délaisse en partie la préposition *en* (issue du latin *in*) au profit d'une autre, *dans*. » (B. Fagard et B. Combettes, 2013 : 93)

3.3. Mise à l'épreuve de l'hypothèse de Darmesteter

On se propose d'éprouver sur le corpus Presto l'hypothèse d'A. Darmesteter en examinant les implications qu'on peut en tirer.

- la première [I₁] est formulée par l'auteur lui-même (cf. *supra*) : « Les premiers emplois [de *dans*] ont dû être ceux où *dans* était suivi de l'article *le* et d'un mot commençant par une consonne, ou de l'article *les*. » ;
- seconde implication [I₂] : on devrait observer une prévalence des articles définis au voisinage¹⁶ de *dans* à l'inverse des autres sous-catégories de déterminants.

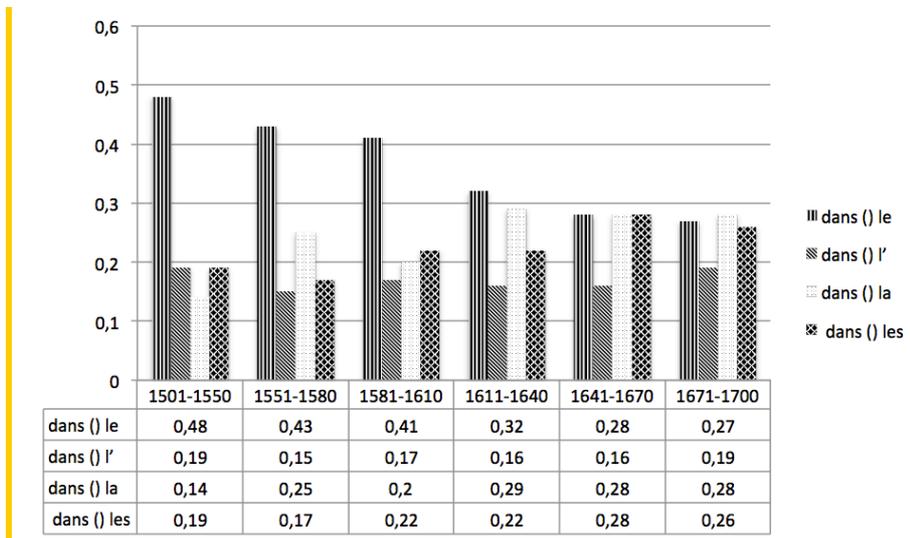
Dans le cadre de cet article et des limites de caractères qui nous sont imparties, nous n'examinerons que la première implication. Nous donnerons sous forme de résumé, en note, les résultats obtenus pour l'étude de la seconde.

Dans la mesure où notre étude se focalise désormais sur une période restreinte de notre corpus (correspondant en gros à la « naissance » de *dans*), nous avons travaillé sur le sous-corpus Presto¹⁵⁰¹⁻¹⁷⁰⁰, que nous avons lui-même scindé en six sous-corpus¹⁷ :

¹⁶ Par « voisinage de *dans* » nous entendons la position qui suit immédiatement la préposition (*dans le cas*) et la seconde position si la première est remplie par le (pré)déterminant « tout » (*dans tous les cas*). Dans Presto, la première occurrence de ce dernier type apparaît en 1607.

¹⁷ Sans pouvoir développer ce point, on observera que nous travaillons désormais sur des « sous-corpus » et non sur des « parties » comme précédemment. Il importe de distinguer soigneusement « partie » et « sous-corpus » en textométrie. Une partie est issue d'une *partition* d'un corpus (voir note 5), de sorte que chaque « partie » implique les autres (la taille de toutes les parties étant égale à la taille du corpus). En revanche, un sous-corpus est le résultat d'un prélèvement opéré sur un corpus de départ sans impliquer d'autres sous-corpus. Si nous avons choisi de travailler ici sur des sous-corpus, c'est qu'il ne s'agit plus pour nous de comparer la distribution d'une forme *i* dans une *partie j* avec sa distribution dans le corpus entier, mais de

le premier couvre une période de cinquante ans¹⁸ (1501-1550), les cinq autres une période de trente ans. Nous avons ensuite calculé, dans chacun de ces sous-corpus, quelle était la fréquence relative¹⁹ de chaque forme de l'article défini au voisinage de *dans*.



Fréquences relatives des formes de l'article défini au voisinage immédiat de *dans* ; corpus Presto¹⁵⁰¹⁻¹⁷⁰⁰ divisé en 6 tranches temporelles

Entre 1501 et 1610, la fréquence relative de la forme *le* domine nettement celle des trois autres formes. Cette prévalence va cependant décroissant et on arrive, pour la période 1641-1700, à une situation de quasi-égalité entre les fréquences affectées à *le*, *la*, *les*. A aucun moment en revanche, la fréquence de *les* ne se détache clairement de celle de *la*, l'une devant l'autre tour à tour, et cela

comparer, au sein d'une tranche chronologique (= *sous-corpus*), la distribution d'une forme *i* (= forme de l'article défini) dans le voisinage d'un pivot (*dans*) par rapport à sa distribution dans le reste de cette même tranche quelle que soit sa position sur l'axe syntagmatique.

¹⁸ Eu égard au nombre plus restreint de textes disponibles pour cette période.

¹⁹ Cette fréquence correspond au quotient d'une fraction prenant pour numérateur la fréquence (absolue) de l'une des formes de l'article défini dans le voisinage de *dans* et pour dénominateur le nombre total d'apparitions de l'article défini (dans la tranche considérée) au voisinage de *dans*.

jusqu'en 1641-1670. Un tel constat pour la forme *les* déjoue clairement une partie des prévisions d'A. Darmesteter.

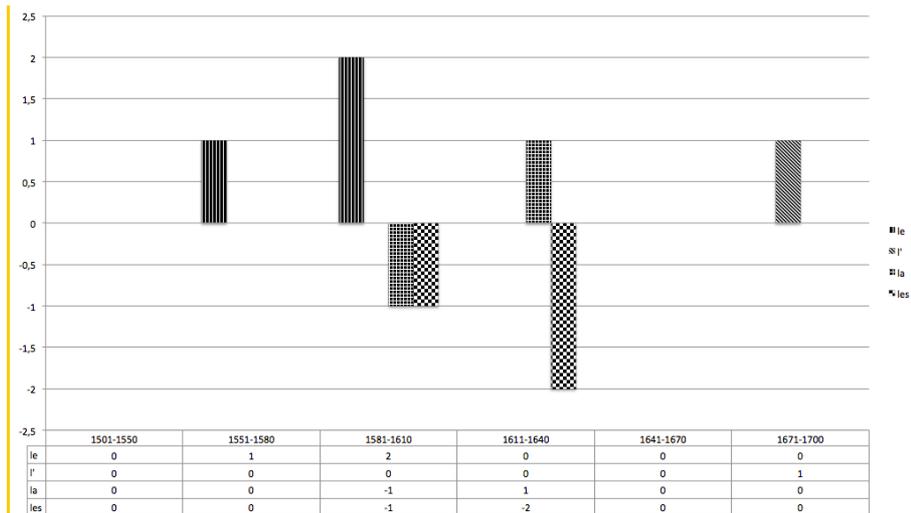
Est-ce à dire que l'on comptera parmi les arguments favorables à la thèse de ce dernier la seule, mais nette, prévalence de *le* – au vu des fréquences relatives – au voisinage de *dans* ? Notre réponse est négative et voici pourquoi. On peut formuler l'hypothèse que, dans chacun des six sous-corpus distingués *supra* (Presto¹⁵⁰¹⁻¹⁵⁵⁰, Presto¹⁵⁵¹⁻¹⁵⁸⁰, Presto¹⁵⁸¹⁻¹⁶¹⁰, etc.), la prévalence de *le* observée au voisinage de *dans* n'est peut-être simplement que le reflet de la distribution de cette même forme dans tout le sous-corpus²⁰.

Pour vérifier une telle hypothèse (qui conteste à *dans* toute combinatoire singulière avec *le*, *contra* A. Darmesteter), nous avons recouru à un calcul des cooccurrences afin de déterminer si la distribution des formes de l'article défini présente ou non une déformation statistiquement significative au voisinage de *dans* par rapport à sa distribution dans l'ensemble du corpus.

Pour chacun des sous-corpus considérés, les paramètres pris en compte ont été :

- f_{ij} = sous-fréquence de la forme de l'article défini (*le|la|les|l'*) au voisinage de *dans* ;
- F_i = fréquence de la forme (*le|la|les|l'*) de l'article défini dans le sous-corpus;
- t_j = sous-fréquence de l'article défini au voisinage de *dans* ;
- T = fréquence de l'article défini dans le sous-corpus;

²⁰ Voir note 14.



Calculs de cooccurrence appliqués au voisinage de la préposition *dans* au sein de cinq sous-corpus constitués dans Presto¹⁵⁵¹⁻¹⁶⁷⁰. Indices de cooccurrence obtenus pour les diverses formes de l'article défini

Tous les scores calculés évoluent dans une zone située en deçà du seuil de significativité statistique de valeur absolue 3 : signe que la distribution des formes de l'article défini au voisinage de *dans* ne présente en réalité aucune déformation statistiquement remarquable et donc qu'elle est à l'image de sa distribution dans le reste du corpus, quelle que soit la position syntaxique du SN dans la composition duquel il entre.

L'étude des données fréquentielles concernant la distribution des formes de l'article défini au voisinage de *dans* nous conduit donc à conclure que :

- contrairement à ce que prévoit l'hypothèse d'A. Darmesteter, la forme *les* n'est pas prévalente dans cette position au regard des trois autres formes relevant de cette catégorie ;
- les distributions des formes de l'article défini au voisinage de *dans* reflètent leurs distributions dans le reste du corpus. Autrement dit la prévalence fréquentielle de *le* dans le voisinage de *dans* jusqu'en 1640 est purement et simplement à l'image de sa prévalence pour la même période dans le reste du corpus et ne constitue pas une caractéristique propre à *dans*.

Tous ces éléments conduisent à reconsidérer l'hypothèse d'A. Darmesteter²¹. La ruine de la « preuve » quantitative entraîne-t-elle avec elle tout l'édifice? Notre point de vue n'est pas si radical et nous plaiderons en conclusion pour un « aménagement » du premier volet de la thèse. Mais auparavant, il convient de montrer que nos outils peuvent aussi apporter de nouveaux éclairages sur l'environnement distributionnel de *dans* au XVI^e s.

4. CONSTRUCTION D'UNE HYPOTHESE ALTERNATIVE

4.1. Point de départ : exploration statistique de la combinatoire *amont* de la préposition *dans* au XVI^e s.

La voie que nous nous proposons d'explorer a été ouverte par B. Fagard et B. Combettes (2013) qui travaillaient sur des données extraites de FRANTEXT et ne disposaient pas pour leur étude de corpus étiqueté et lemmatisé.

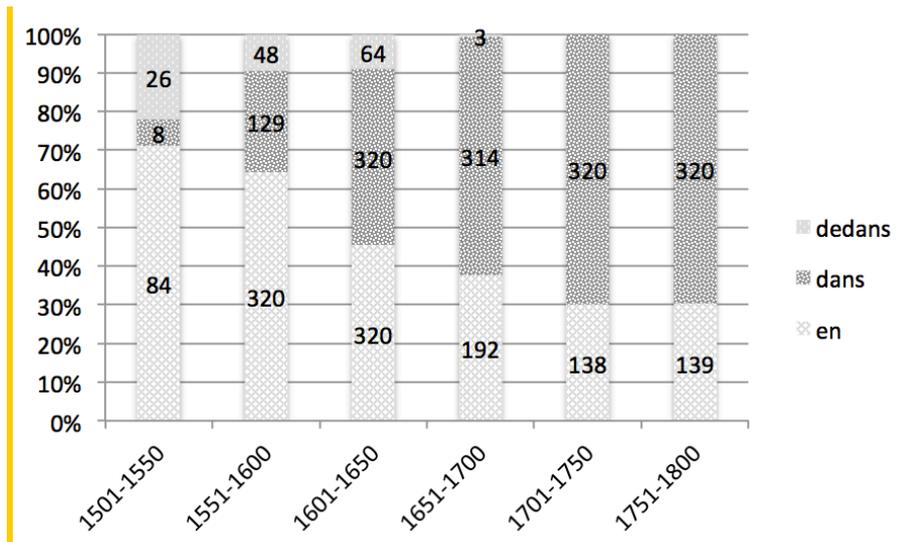
Désireux d'analyser « l'impact du contexte sémantique plus large sur l'évolution de *en* et d'autres prépositions, entre XVI^e et XVII^e siècles » (102), les auteurs choisissent d'étudier la combinatoire aval des verbes *entrer* (« intransitif de mouvement »), *jeter*, *lancer* (« mouvement causé »). Le choix de ces verbes est de leur propre aveu « en partie arbitraire, et doit être vu comme un point de départ pour une réflexion sur le fonctionnement de *en*, *dans* et *dedans* » (*ibid*). La conclusion à laquelle ils aboutissent est que dans ces contextes, on observerait pour l'essentiel un phénomène de remplacement de *en* par *dans*, accompagné d'un changement notable dans l'emploi du déterminant devant le régime nominal. « [On observe une] opposition frappante entre XVI^e et XVII^e siècles sur plusieurs points. Le premier est que *dans* s'impose très rapidement comme préposition introduisant le complément des verbes *entrer* et *jeter* aux dépens de *en*. Le second est que *en* se spécialise dans l'introduction de N nus, i.e. sans déterminant. » (*ibid* : 103)

²¹ L'examen de l'implication 2 est tout aussi défavorable à la thèse examinée. Ce n'est qu'au cours du XVII^e s. en effet que l'article défini s'avère une sous-catégorie de déterminants sur-utilisée au voisinage de *dans* par rapport à sa distribution dans le corpus entier. Le caractère très tardif du phénomène est incompatible avec le raisonnement tenu par A. Darmesteter.

Dans le but à la fois d'affiner certains aspects de cette étude et de consolider (voire d'élargir) la gamme de verbes recteurs dans le régime desquels on peut observer au XVI^e s. une préférence significative pour les SP à tête *dans*, nous avons accompli un calcul des cooccurrences pour déterminer i) quelles catégories morphosyntaxiques et ii) quels lemmes s'avéraient sur-représentés dans le voisinage *amont* de *dans* au sein de Presto¹⁵⁰¹⁻¹⁶⁰⁰. Le résultat est sans appel : la catégorie morphosyntaxique préférée pour ce cotexte de cooccurrence est celle des verbes ; les cinq premiers lemmes verbaux qui apparaissent par ordre de préférence décroissant sont : *entrer, mettre, jeter, enfermer, cacher*²².

Cette liste de verbes préférés étant arrêtée, on peut renverser le calcul et faire de ces verbes le pivot d'une analyse cooccurentielle pour étudier les préférences combinatoires qu'ils manifestent à l'égard de leur régime. Le diagramme ci-dessous présente les résultats des calculs opérés sur le cotexte droit (fenêtre de deux mots) de ces verbes saisis ensemble dans un seul pivot complexe, et cela pour la période temporelle 1501-1800. Le recours à ce type de diagramme permet de visualiser, pour chaque tranche temporelle, la part relative (exprimée en pourcentage) occupée par l'indice de cooccurrence affecté à chacune des trois prépositions (indice exprimé dans chaque segment de la barre) sur la somme totale de ces trois indices.

²² Si l'on se réfère au classement des verbes notamment de *déplacement* vs *mouvement* proposé par M. Aurnague (par ex. 2012), *entrer* ressortit aux verbes de déplacement, *jeter* aux verbes de déplacement causé, *mettre* aux verbes de changement d'état, *enfermer* et *cacher* aux verbes d'inclusion sans déplacement.



Evolution de la part occupée (en %) par chacun des indices de cooccurrence affectés aux prépositions *en*, *dans*, *dedans* à la suite du pivot complexe. [lemma=" ENTRER|METTRE|JETER|ENFERMER|CACHER"]. Contexte de cooccurrence de 2 mots à droite du pivot ; corpus Presto¹⁵⁰¹⁻¹⁸⁰⁰, tranches de 50 ans

L'évolution des pourcentages occupés par les scores de spécificité de *dans* et de *en* est frappante: tandis que le premier ne cesse de croître entre 1501 et 1750, le second ne cesse de baisser jusqu'à la même date. *Dans* occupe seul (et définitivement) le haut du podium des prépositions préférées du pivot à partir de la tranche temporelle 1651-1700. Quant à *dedans* il disparaît à l'aube du XVIII^e s. de la liste des cooccurrents préférés.

Pour affiner notre compréhension de l'évolution des distributions de *dans* et de *en* au voisinage des verbes sélectionnés, nous avons fait porter notre enquête sur l'identification des cooccurrents nominaux les plus spécifiques situés dans le régime des verbes *entrer* et *mettre*²³. Faute de place, nous présentons ci-dessous les listes des cooccurrents

²³ La fréquence moindre des trois autres verbes dans notre corpus n'a pas permis une application satisfaisante de nos outils.

nominaux les plus spécifiques pour le seul verbe *entrer*²⁴ (l'ordre des mots dans les cellules correspond à leur ordre hiérarchique de spécificité décroissante).

	XVI	XVII	XVIII	XIX
Pivot	Liste des 5 premiers cooccurrents nominaux (lemmes) les plus spécifiques ²⁵ pour chaque pivot			
Pivot 1 <i>ENTRER</i> en N	<i>possession,</i> <i>propos,</i> <i>dispute,</i> <i>matière,</i> <i>colère</i>	<i>possession,</i> <i>matière,</i> <i>comparaison,</i> <i>religion,</i> <i>désespoir</i>	<i>lice,</i> <i>conférence,</i> <i>concurrence,</i> <i>négociation,</i> <i>matière</i>	<i>action, lice,</i> <i>partage,</i> <i>possession,</i> <i>ligne</i>
Pivot 2 <i>ENTRER</i> en dét. N	<i>chambre,</i> <i>maison,</i> <i>église,</i> <i>entendement,</i> <i>ville</i>	<i>salle,</i> <i>chambre,</i> <i>pays, cabinet,</i> <i>possession</i>	∅	∅
Pivot 3 <i>ENTRER</i> dans dét. N	<i>pays,</i> <i>chambre</i>	<i>chambre,</i> <i>ville, port,</i> <i>détail,</i> <i>sentiment</i>	<i>détail,</i> <i>chambre,</i> <i>ville, maison,</i> <i>composition</i>	<i>détail,</i> <i>chambre</i> ²⁶ , <i>composition,</i> <i>maison, voie</i>
Liste des cooccurrents nominaux spécifiques PARTAGÉS ²⁷ par deux pivots dans une même tranche temporelle				

²⁴ Les conclusions que nous allons dégager sont parfaitement cohérentes avec celles que nous avons tirées de l'examen du profil cooccurrentiel de *mettre*.

²⁵ Répartis dans au moins *trois* textes distincts de façon à corriger les distorsions introduites par les cas de fréquences élevées, dans un seul texte, de mots par ailleurs peu usités dans le reste du corpus.

²⁶ La prévalence pendant cinq siècles du N *chambre* après *dans* a été remarquée par P. Blumenthal (2011 : 293) sur un autre corpus.

²⁷ Les lemmes partagés sont prélevés dans l'ensemble des cooccurrents les plus spécifiques des pivots concernés et non seulement parmi les cinq premiers. D'où la possibilité pour certains d'entre eux de ne pas figurer dans les lignes 3 à 5 du tableau.

Pivots 1 & 3	∅	<i>composition</i> <i>défiance</i>	<i>composition</i>	<i>combinaison</i>
Pivots 2 & 3	<i>chambre</i>	<i>chambre,</i> <i>salle, cabinet,</i> <i>pays, esprit,</i> <i>maison,</i> <i>église,</i> <i>composition,</i> <i>place,</i> <i>royaume</i>	∅	∅

Listes des lemmes nominaux les plus spécifiques situés dans le régime du lemme verbal *entrer*, corpus Presto ^{-ETENDU_1501_1900}

Les éléments les plus saillants qui ressortent de l'examen de ce tableau sont :

- Au XVI^e s, lorsque *dans* est intégré dans la combinatoire de *entrer*,
- ce verbe présente déjà une combinatoire nominale étoffée avec *en* ;
- la combinatoire nominale spécifique qu'il sélectionne avec *dans* réunit des noms dénotant des réalités dotées d'extension matérielle ou physique (*pays, chambre*) qui jouent contextuellement le rôle de site de repérage spatial d'une cible;
- il existe d'emblée une combinatoire nominale spécifique partagée pour les constructions *ENTRER en dét. N* et *ENTRER dans dét. N*.

Le XVII^e s. constitue un siècle charnière. On y observe en effet:

- une combinatoire nominale étoffée pour *chacun* des trois pivots, ce qui ne sera plus le cas ensuite ; concernant *dans* plus particulièrement, apparaissent dans sa distribution spécifique, outre des noms d'entités spatiales, des noms abstraits dénotant des réalités sans extension matérielle ou physique (*détail, sentiment*).
- Les combinaisons nominales *partagées* entre deux voire trois pivots s'avèrent remarquablement étoffées.

Au delà, au XVIII^e s. et au XIX^e s., la construction *entrer en dét. N* s'éteint rapidement. Parallèlement, la combinatoire la plus spécifique de *entrer dans dét. N* continue à présenter une série hybride de noms

dénotant des réalités contextuellement dotées – ou non – d’extension matérielle ou physique²⁸. Enfin, la structure *entrer en N* sélectionne désormais spécifiquement des noms (contextuellement) abstraits.

Ces observations mises en faisceaux permettent de décrire un processus global dans lequel nous distinguerons deux étapes :

XVI^e-XVII^e s. : vers une ouverture maximale de la combinatoire nominale de *ENTRER (en + dans)*.

L’arrivée de *dans* au sein de la combinatoire de *entrer* conduit rapidement (XVII^e s.) à la cohabitation de trois combinatoires nominales qu’on peut répartir en termes de « polarité » concrète/ abstraite comme suit :

- deux combinatoires à polarité nominale hybride concrète-abstraite (noms dénotant contextuellement des réalités dotées ou non d’extension matérielle ou physique) : *entrer (en + dans) dét. N*

- une combinatoire à polarité nominale abstraite : *entrer en N*

Dès le XVI^e s., la construction *entrer dans* conquiert une partie de sa combinatoire nominale spatiale sur celle de *entrer en dét. N*. Au siècle suivant, *dans* s’approprie aussi une partie de la combinatoire abstraite de cette même construction. En ce sens, on peut dire qu’il y a eu un remplacement de *en* par *dans* devant certains noms.

Le résultat de ce premier processus conduit à une situation « exceptionnelle » au XVII^e s. où les constructions *entrer en* et *entrer dans* partagent une combinatoire nominale spécifique particulièrement substantielle.

XVIII^e s. - XIX^e s.. Processus de reconfiguration et de spécialisation de la combinatoire nominale de *entrer (dans + en)*.

A l’aube du XVIII^e s., on observe une raréfaction puis une disparition rapide de la combinatoire nominale de la construction *entrer en dét. N*. La combinatoire partagée entre les pivots 2 et 3 connaît le même destin.

²⁸ On observera au passage que trois des cooccurrents nominaux préférés de *entrer en dét.* au XVI^e s. (*chambre, maison, ville*) sont passés sous la coupe de *entrer dans dét.*

Il s'opère alors une spécialisation des deux constructions restantes :

- *entrer dans* dotée d'une combinatoire à polarité hybride s'avère apte i) à opérer des localisations spatiales-concrètes d'une cible dans un site au terme d'un déplacement ; ii) à exprimer divers rapports plus abstraits (*entrer dans la composition de*²⁹, ...)
- *entrer en N* dotée d'une combinatoire nominale abstraite non actualisée, tend à se figer en des locutions verbales (*entrer en lice, entrer en matière, ...*)

4.2. Vers la formulation d'une nouvelle hypothèse

Si l'on revient à l'argumentation présentée dans notre première partie et relative à la question des amalgames issus de *en*, il nous paraît incontestable que les locuteurs du XVI^e s. ont été placés face à un déséquilibre du système engendré par la substitution des formes *au, aux* aux amalgames morphologiquement issus de **en le / *en les*. En d'autres termes, nous ne remettons pas en cause l'hypothèse selon laquelle la cause première (le *primum movens* en quelque sorte) du changement linguistique profond que marqua l'ascension fulgurante des emplois de *dans* en français préclassique fut cette « surcharge » due à l'ambiguïté de *au, aux* qu'évoque G. Gougenheim (1951 : 182 ; cf. *supra* 1.2.1.). Là où nous divergeons d'avec A. Darmesteter et ses successeurs, c'est sur le profil distributionnel qui fut celui de *dans* en français préclassique. Rien dans notre corpus ne permet de considérer que cette préposition se serait d'abord préférentiellement substituée à *en* devant les formes *le* et *les* de l'article défini. Tout porte à croire au contraire que la diffusion de *dans* s'est d'emblée accomplie devant toutes les formes de déterminants³⁰. En revanche, nos outils font

²⁹ Un de nos relecteurs a fait justement observer que si coexistent en français moderne les séquences *entrer en composition* et *entrer dans la composition*, la complémentation attendue est distincte sur les plans syntaxique et constructionnel: *entrer en composition avec X* vs *entrer dans la composition de X*. L'exploration du cotexte droit permet du même coup d'affiner les hypothèses sur le cotexte gauche de la préposition.

³⁰ Dans la thèse d'A. Darmesteter, l'ouverture de *dans* à tous les déterminants ne s'accomplit que *dans un deuxième temps* : « La langue ne pouvait se résoudre à n'employer *dans* qu'avec l'article *le* ou *les*, exactement dans les cas où il représentait *ou* et *es*. C'aurait été imposer à ses habitudes un formalisme et une rigueur inconnue de l'esprit populaire. Celui-ci (...) étendit l'emploi de *dans* à toutes les expressions où le substantif est déterminé : *dans la maison, dans cet état, dans toute affaire, dans ces circonstances*, etc. » (*ibid.*, 186). D'après nous, c'est

apparaître une préférence des SP ayant pour tête *dans*, pour la position valencielle de certains verbes marquant notamment le mouvement, le déplacement, le changement d'état.

Ajoutons pour finir que très vite (dès la première moitié du XVII^e s.), *dans* s'est affranchi des conditions de sa naissance pour commencer sa vie propre – tout comme *en*³¹ d'ailleurs. Mais c'est là une autre histoire...

immédiatement que cette extension se produit.

³¹ Ce que F. Brunot (1967 : 469) avait parfaitement observé : « *En* commence une autre histoire, du jour où cette préposition cesse de se contracter avec l'article en *ou* et *es*, et où *dans* entre en concurrence avec lui. »

Références

- AURNAGUE M. (2012), « De l'espace à l'aspect : les bases ontologiques des procès de déplacement », *Corela* [En ligne], HS-12 | 2012, mis en ligne le 04 avril 2013, consulté le 30 septembre 2016. URL : <http://corela.revues.org/2846> ; DOI : 10.4000/corela.2846
- BLUMENTHAL P. (2011), « Le figement. Du XVII^e s. au français contemporain », in J.-C. Anscombre, S. Mejri (éds), *La parole entravée*, Paris : H. Champion, 283-302.
- BRUNOT F. (1967), *Histoire de la langue française. Des origines jusqu'en 1900*. Tome II: *le seizième siècle*, Paris : A. Colin.
- DARMESTER 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 » ? », in O. Bonami, S. Prévost, M. Charolles, J. François et C. Schnedeker (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 ». in J. François, E. Gilbert, C. Guimier et M. Krause (dir.), *Autour de la préposition: position, valeurs, statut et catégories apparentées à travers les langues*, *Bibliothèque de Syntaxe et Sémantique*, Caen : PUC, 225-236.
- FAGARD B. & COMBETTES B. (2013), « De en à dans, un simple remplacement ? Une étude diachronique », *Langue Française* 178, 95-119.
- GODEFROY F. (1891-1902), *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*, Paris.
- GOUGENHEIM G. (1945), « Les prépositions « en » et « dans » dans les premières œuvres de Ronsard ». in *Etudes de grammaire et de vocabulaire français, réunies sur l'initiative de ses collègues et amis pour son soixante-dixième anniversaire*, Paris : Picard, 55-65.
- GOUGENHEIM G. (1950, 1970), « Valeur fonctionnelle et valeur intrinsèque de la préposition « en » en français moderne », in *Etudes de grammaire et de vocabulaire français, réunies sur l'initiative de ses collègues et amis pour son soixante-dixième anniversaire*, Paris : Picard, 55-65.

- GOUGENHEIM G. (1951, 1974), *Grammaire de la langue française du seizième siècle*, Paris : Picard.
- GOUGENHEIM G. et al. (1956 [1964]), *L'élaboration du français fondamental*, Paris : Didier.
- HUGUET E. (1925-1967), *Le dictionnaire de la langue française du seizième siècle*, 7 volumes, Paris : Champion / Didier.
- LABBE C. et LABBE D. (1994), « Que mesure la spécificité du vocabulaire ? », Grenoble, CERAT. Repris dans : *Lexicometrica* 3, 2001.
- LAFON P. (1980), « Sur la variabilité de la fréquence des formes dans un corpus », *Mots* 1, 127-165
- LAFON P. (1984), *Dépouillements et statistiques en lexicométrie*, Genève/Paris : Slatkine Champion.
- LARDON S., THOMINE, M.-C. (2009), *Grammaire du français de la Renaissance*, Paris : Garnier.
- LEBART L. & SALEM A. (1994), *Statistique Textuelle*, Paris : Dunod.
- REUTENAUER C. (2002), *Vers un traitement automatique de la néosémie*, Thèse nouveau régime, Université de Lorraine.
- SALEM A. (1988), « Approches du temps lexical. Statistique textuelle et séries chronologiques », *Mots* 17, 105-143.
- TERREAUX L. (1968), *Ronsard correcteur de son œuvre. Les variantes des Odes et des deux premiers livres des Amours*, Genève : Droz.
- VAGUER C. (2008), « Classement syntaxique des prépositions simples du français », *Langue française* 157, 20-36.
- VIGIER D. (2015), « Les prépositions *en*, *dans* et *dedans* au XVIe s. Approche statistique et combinatoire », *Le Français moderne*, vol. 2, 230-247.