

Étude de l'évolution sémantique des prépositions *à, en, dans, dedans* du français. Quel(s) apport(s) d'une périodisation automatique ?

Diwersy Sascha (1), Falaise Achille (2) & Vigier Denis (3)

(1) Montpellier 3, UMR Praxiling

(2) ENS Lyon, UMR ICAR

(3) Lyon 2, UMR ICAR

sascha.diwery@univ-montp3.fr, achille.falaise@ens-lyon.fr, denis.vigier@univ-lyon2.fr

Le corpus diachronique du français PRESTO, constitué dans le cadre du projet franco-allemand du même nom (<http://presto.ens-lyon.fr>), échantillonné, contrôlé et équilibré (critères chronologiques, d'auteurs et de genres discursifs), permet aujourd'hui d'étudier en diachronie longue (1509-1944) l'évolution du profil cooccurentiel des prépositions du français. C'est ce que nous proposons ici de faire pour les prépositions *à, en, dans, dedans*, en nous posant plus spécifiquement la question de la périodisation des phénomènes cooccurentiels sur lesquels nous travaillons.

Rappelons d'abord qu'on peut tenir 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 pour figurer finalement, en français contemporain, au cinquième rang en termes de fréquence d'emploi derrière *de, à, en* et *pour*¹. Or si l'on examine au moyen du calcul des cooccurrences sur TXM (calcul des spécificités de P. Lafon) et sur PrimeStat (calcul du log likelihood) quels sont les accompagnateurs nominaux préférés de *dans* lors de ses premiers emplois, on observe que les noms qui figurent parmi ses collocatifs les plus spécifiques (au sens statistique) dénotent tous une réalité dotée d'une extension matérielle ou physique : lieux géographiques construits de la vie quotidienne (*château, ville, maison, salon, ...*) ou naturels (*mer, bois, plaine, ...*), noms désignant des parties et des productions du corps humain (*entrailles, sang, sein*). Ces réalités constituent, en discours, des sites de repérage pour une cible, au terme (1) ou non (2) d'un déplacement dans l'espace :

(1) *Le president entra dans la chambre et trouva sa femme et nicolas couchez ensemble.* (1550, M. de Navarre, *L'Heptameron*)

(2) *Ce pendant qu'on est en ce monde, On est dans une mer profonde (...)* (1587, Pierre de l'Estoile, *Registre-journal du regne de Henri III*)

Ce profil cooccurentiel de *dans* qui s'affirme lors de ses premiers emplois en français préclassique fait de cette préposition nouvellement arrivée **une rivale distributionnelle** de trois autres prépositions beaucoup plus anciennes :

- *en* qui, issue du latin, couvre un ensemble d'emploi qui se partagent de manière équilibrée entre régimes nominaux déterminés et non déterminés et qui s'avère apte elle aussi à opérer des localisations de cible dans les limites d'un site ; par ex.

(3) *Quant le mary veid qu'il en avoit bien fait son debvoir, entra en la chambre et le mercia de la peyne qu'il en avoit prinse* (1550, M. de Navarre, *L'Heptameron*)

- (4) *Mais tousjours demouroit en la nef entre les femmes* (1532, F. Rabelais, *Pantagruel*)
- *dedans*, qui semble se distinguer de *en* par sa préférence pour des contextes où la cible est entièrement recouverte par le site :
- (5) *Or on ne peut cognoistre cela aux poissons, car on ne peut sçavoir leur aage, d'autant qu'ils vivent dedans l'eau.* (1556, B.-G. Gelli, *Les discours fantastiques de Julien Tonnelier*)
- Quant à *à*, on sait que les formes amalgamées *au*, *aux* qui en étaient issues pouvaient être recrutées aussi bien dans les contextes où cette préposition se trouvait au contact des formes *le*, *les* de l'article défini, que dans les contextes où les formes issues de la fusion de *en* et de *le/les*² étaient déficientes – *au*, *aux* entrant alors dans le paradigme de *en* :
- (6) *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 S. Lardon & M.-C Thomine, *ibid.* : 393)

En d'autres termes, tout porte à croire que les « destins » cooccurrentiels de *à*, *en*, *dans*, *dedans* ont été – du moins pendant une certaine période après 1550 – peu ou prou liés. Nous nous proposons donc dans cette communication d'étudier **l'évolution sémantique de ces quatre prépositions entre le XVI^e e s. et le XX^e e s.**, notre hypothèse étant que l'irruption de *dans* sur la scène des prépositions a profondément recomposé le profil distributionnel des trois autres.

Après avoir brossé le contexte au sein duquel *dans* paraît vers 1550, nous présenterons les résultats d'une étude comparée que nous avons accomplie sur le corpus PRESTO 1501-1950 et portant de l'évolution des profils cooccurrentiels de ces quatre prépositions. Cette étude s'adosse (i) d'une part à une série de calculs de spécificités conduits sur une partition du corpus diachronique PRESTO (tranches de cinquante années découpées arbitrairement en s'appuyant sur les frontières de (demi-)siècles), (ii) d'autre part sur une suite de calculs de cooccurrence lexicale conduits sur des sous-corpus et prenant pour pivot successivement *dans*, *en*, *dedans*, *à*. Les résultats de ces calculs ont été analysés manuellement, en vue d'identifier pour chacune des prépositions d'éventuelles zones temporelles où apparaissent des recompositions particulièrement prononcées de leur profil combinatoire.

Puis nous nous tournerons vers la méthode de *classification ascendante hiérarchique par contiguités* (CAHC, de l'anglais *Variability based neighbour clustering* – cf. Gries & Hilpert 2008 ; 2012 ; Hilpert 2013 : 32-45), qui vise avant tout à produire une périodisation automatique. Cette méthode peut par exemple s'appliquer aux paradigmes cooccurrentiels de *en*, définis en termes des collocatifs nominaux spécifiques³ relevés, dans une fenêtre de trois mots à droite, et par tranches de 25 années, dans le corpus PRESTO étendu (1501-2010). On peut alors constater (figure 1), que les emplois de cette préposition font l'objet d'une reconfiguration fondamentale, qui est entamée pendant la première moitié du XVII^e siècle et qui va depuis dans le sens d'une spécialisation (cf. Diwersy, Falaise, Lay et Souvay 2017).

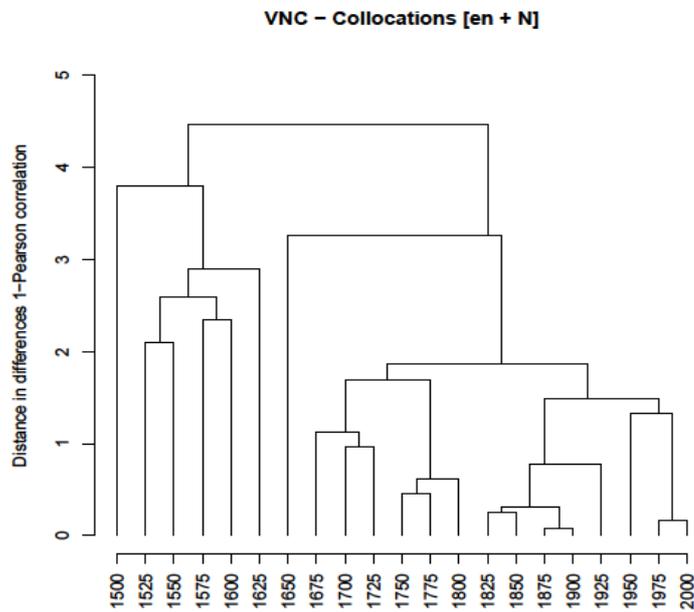


Figure 1 : CAHC appliquée aux paradigmes collocationnels formés par les noms constituant un SN régi par *en* (corpus PRESTO étendu, tranches de 25 années).

Notre objectif sera de comparer les résultats obtenus par les deux méthodes d'analyse ci-dessus, la première, « hybride » (« analyse des collocations », au sens de K. Heylen et A. Bertels (2016 : 53)), en ceci qu'elle combine identification automatique et statistique des indices contextuels et traitement manuel des résultats, et la seconde, entièrement automatique (méthodes d'« analyse distributionnelles »).

Références bibliographiques

- 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.
- DIWERSY S., FALAISE A., LAY M.-H., SOUVAY G. (2017). Ressources et méthodes pour l'analyse diachronique. *Langages* 206, à paraître.
- 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.
- 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.
- GRIES S.T. & HILPERT M. (2008), « The identification of stages in diachronic data: variability-based neighbour clustering », *Corpora* 3 (1), 59-81.
- GRIES S.T. & HILPERT M. (2012), « Variability-based neighbor clustering: a bottom-up approach to periodization in historical linguistics », in T. Nevalainen & E. Traugott (eds), *The Oxford handbook of the history of English*, Oxford : Oxford University Press, 134-144.
- HEYLEN, K., BERTELS, A. (2016), « Sémantique distributionnelle en linguistique de corpus », *Langages*, 201, 51-64
- HILPERT M. (2013), *Constructional Change in English: Developments in Allomorphy, Word formation, and Syntax*, Cambridge : Cambridge University Press.

¹ Nous avons étudié ailleurs les raisons de cette surprenante fortune de *dans* en français.

² El, on, ou, .../ els, es, ès, ez, ...

³ Ont été retenus comme spécifiques tous les collocatifs nominaux dont le score d'association *log-likelihood* était égal ou supérieur à 10,83 (valeur-p < 0,001).