

C'est très théâtre, c'est-à-dire très faux. *Les origines et le développement de la construction* *[Adverbe_{degré} + Nom]*

PETER LAUWERS

Universiteit Gent (Université de Gand)

(Received October 2017; revised July 2018)

RÉSUMÉ

Dans cette contribution, nous nous intéressons à l'origine et au développement de la construction [adverbe de degré + nom]: *C'est très théâtre, c'est-à-dire très faux*. De nos jours, cette construction se présente comme un puissant outil de recatégorisation contextuelle pour exprimer un rapport de ressemblance basé sur une analogie avec un concept nominal. Sur la base d'une recherche de corpus outillée, nous montrons que la construction-hôte [ADV_{degré} + ADJ] a pu s'ouvrir à la catégorie nominale à partir d'un petit noyau dur de noms humains, en partie qualitatifs, tirant profit de certaines propriétés structurelles du français de l'époque. Par la suite, la construction s'est diversifiée vers d'autres classes sémantiques (noms inanimés, noms propres, etc.) et a progressivement attiré une gamme plus diversifiée de types lexicaux dans son orbite. Elle n'a cessé de gagner en productivité, tout en réduisant son fonds lexical d'origine à la portion congrue.

I. INTRODUCTION: UNE CONSTRUCTION PROVOQUANT DES EFFETS DE RECATÉGORISATION EN DISCOURS

Le problème de la transcatégorialité des lexèmes sans marquage morphologique ouvert constitue un réel défi à la grammaire. Ses multiples avatars ont été traités dans un nombre considérable d'études qui souvent partagent le champ en deux domaines: la *conversion* et la *distorsion catégorielle* (Kerleroux, 1996, 1999) ou *coercition* (Lauwers et Willems, 2011; Lauwers, 2014a). Ce clivage subsume en réalité un ensemble d'oppositions qui se situent à différents niveaux:

- le degré de recatégorisation: complet ou partiel
- le niveau d'incidence: morphologie lexicale (double appartenance catégorielle) ou syntaxe
- le statut ontologique du résultat: phénomène de langue ou de discours.

Nous nous pencherons ici sur un phénomène qui relève de la distorsion catégorielle: la recatégorisation en contexte des noms. Il s'avère en effet que le français contemporain dispose de plusieurs constructions (Lauwers, 2014b) dans lesquelles des noms apparaissent occasionnellement dans des contextes syntaxiques

typiques de l'adjectif (qualificatif), donnant lieu à des effets de recatégorisation éphémères, en contexte. L'une de ces constructions, la plus ancienne, fera l'objet de la présente contribution:

- (1) C'est **très théâtre**, c'est-à-dire très faux. (Renard, *Journal*, 21 sept. 1908; cité dans Grevisse et Goosse, 1991: 1462)
- (2) Patrick est **très mère-poule** avec ses pigeons. (Google, site de colombophilie, 31/08/2017)

Comme nous avons déjà analysé en détail la construction en synchronie (Lauwers, 2014b; cf. aussi López Díaz, 2008), nous nous limiterons ici à en résumer l'essentiel, pour passer ensuite à l'analyse diachronique, le véritable enjeu de la présente contribution. La construction qui nous intéresse se caractérise par la présence d'un adverbe de degré ou d'intensité (*très, trop, si*, etc.) suivi d'un nom, accompagné éventuellement d'un modificateur adnominal qui est soumis à des contraintes assez fortes. Sur le plan du sens, elle exprime un rapport de ressemblance – rapport souligné quelquefois par l'emploi évidentiel du verbe *faire* (*ça fait très théâtre*) – entre un support (*ce*) et un apport (*théâtre*). Même s'il n'y a pas de contrainte sémantique absolue, en général apport et support convergent en termes d'animéité (animé/animé ou inanimé/inanimé). Les seules contraintes sont de nature pragmatique. En effet, comme le rapport de ressemblance est basé sur des propriétés associées stéréotypiquement au nom apport, il faut que le nom apport fasse l'objet de stéréotypes sociaux partagés (cf. Goes, 1999: 169), sinon la phrase n'est pas acceptable:

- (3a) Mon père est **très cure**.
- (3b) des conseils **très médecin de famille**
- (3a') ??Julien est **très membre**.
- (3b') ??une table **très chaise**

Les stéréotypes mobilisés varient d'un contexte, voire d'un locuteur à l'autre. Ainsi, dans (1) c'est l'hypocrisie que Jules Renard attribue au petit monde parisien du théâtre, dans d'autres cas cela pourrait être le caractère surjoué (d'un dialogue de film, par exemple). Sur le plan syntaxique, la construction confère un haut degré d'adjectivité à l'item recatégorisé, ce qui a partie liée avec le fait que la distorsion catégorielle opère à l'intérieur du syntagme adjectival, comme en témoigne aussi la liberté syntaxique du syntagme qui en résulte [Adv + N'], qui peut en effet fonctionner comme attribut, apposition ou épithète¹ (voir Lauwers, 2014b).

Ici nous poursuivrons l'enquête en nous interrogeant sur les origines (3.) et le développement de la construction (4.), ce qui revient, globalement, à examiner sa productivité. Mais regardons d'abord quelques aspects méthodologiques de cette recherche (2.).

¹ La problématique du nom épithète (Noailly, 1990) n'intervient pas dans cette étude, même si récemment ce type de construction a donné lieu à des phénomènes de debonding (cf. Amiot & Van Goethem, 2012).

2. MÉTHODE

Nous avons limité notre examen à cinq adverbes de degré dont les préférences catégorielles adjectivales sont assez nettes: *si*, *très* (et sa variante orthographique *tres*), *trop*, *plus* et *assez*. Dès lors, certains adverbes moins sélectifs et à portée plus variable comme *plutôt* ont d'emblée été écartés (p.ex. *je préfère plutôt celui-là*). Ces cinq adverbes illustrent différents types de gradation et pour presque chacun d'entre eux d'autres items auraient pu être choisis: le haut degré (*très*), le haut degré intensif (*si*; cf. aussi *tellement*, *aussi*), le haut degré dépassant une norme (*trop*, cf. aussi *trop peu*), la norme minimale (*assez*), et, enfin, les degrés de comparaison (*plus*, y compris le superlatif *le plus*; cf. aussi *(le) moins*).²

Comme nous sommes uniquement intéressé par les items lexicaux intrinsèquement nominaux, il fallait exclure des comptages tous les **items intrinsèquement bicatégoriels**. Ces noms-adjectifs peuvent être divisés en deux sous-ensembles:

- Items **polysyntaxiques**: même sens, mais coulé dans un autre moule catégoriel
 - un ami/un collègue ami
 - un catholique/la religion catholique
 - le rouge/un pantalon rouge
 - un petit coquin/une petite fille coquine
- Items **+/- homonymiques**: le passage à une autre catégorie va de pair avec un glissement de sens
 - un sorcier/(c'est pas) sorcier
 - une vache/une prof vache
 - un sujet/sujet (à)

Nous avons été sévère dans l'application³ du critère d'exclusion. Ont ainsi été écartés même les noms n'apparaissant qu'assez rarement en emploi adjectival (sans adverbe de degré, bien entendu) ou uniquement en emploi adjectival relationnel (p.ex. *(peuple) pêcheur*). Il n'empêche que, pour la période la plus ancienne, nous avons aussi relevé la liste des lexèmes bicatégoriels, sans pour autant quantifier leurs occurrences. Cette stratégie a payé, car il s'est avéré que les N bicatégoriels ont joué un rôle important dans l'émergence de la construction.

Vu la nature du phénomène, qui concerne les marges du système grammatical, nous nous sommes tourné d'emblée vers le corpus diachronique le plus vaste dont nous disposons à l'heure actuelle, Frantext. Pour obtenir un nombre maximal d'occurrences ainsi qu'une représentativité maximale⁴ pour l'ensemble de la langue

² Les emplois comme adverbe de négation (*plus*) et conjonction (*si*) ont bien sûr été exclus.

³ Pour cela nous nous sommes basé sur nos intuitions linguistiques et quelques ouvrages lexicographiques (essentiellement *Le Petit Robert* et *TLLFi*). Pour les périodes plus anciennes, nous avons également consulté le *DMF*, Huguét (16^e siècle) et, pour les items problématiques, Frantext non catégorisé.

⁴ Notons qu'il existe une certaine tension entre la volonté d'utiliser un corpus très vaste et le principe de la représentativité. En effet, en voulant inclure un maximum de textes, on

écrite, nous avons inclus tous les genres. Nous avons commencé par Frantext catégorisé (1830–1997), remontant progressivement dans le temps, jusqu’à l’année 1700. Puis, afin de nous faire une idée des évolutions antérieures, nous avons choisi au hasard une période antérieure (1560–1619). Il importe de signaler que nous n’avons pas essayé de trouver la toute première attestation de la séquence [Adv_{deg} + N]. Une telle entreprise aurait de toute façon été vaine, dans la mesure où tous les adverbes de degré (*moult*, par exemple) n’ont pas été examinés. Pour la période la plus récente (postérieure à 1830), nous avons pu nous baser sur Frantext catégorisé. Les choses se compliquaient pour les périodes plus anciennes. Face au nombre considérable de données – des centaines de milliers – produites par la recherche non catégorisée, nous avons dû avoir recours à des outils informatiques pour traiter les données brutes. Pour la période 1700–1829 (Frantext non catégorisé), nous avons également commencé par effectuer une recherche lexicale. L’output de ces requêtes a été reporté dans un grand fichier txt qui a ensuite été étiqueté moyennant Tree Tagger 1.0., disponible sur la plate-forme Sketch Engine (Kilgarriff *et al.*, 2014). Puis, le corpus étiqueté a été interrogé pour repérer les combinaisons *plus/si/assez/très/trop* + Nom. Enfin, pour les périodes plus anciennes, il n’aurait plus été raisonnable d’utiliser Tree Tagger. Comme il fallait parcourir manuellement des milliers d’occurrences, nous nous sommes vu contraint de limiter notre investigation à une période plus courte (1560–1619), ce qui a entraîné un ‘trou’ de 80 ans (entre 1620 et 1699). Pour faciliter le parcours manuel des données, nous avons tout de même procédé au classement alphabétique du terme subséquent (faisant suite à *très*, *si*, etc.) par un concordancier, AntConc.

Si une telle démarche était la seule possible face à un nombre aussi considérable d’attestations, elle a le désavantage de donner lieu à quelques asymétries (en plus du trou signalé ci-dessus). Ainsi, on ne saurait comparer tels quels les résultats d’une recherche lexicale, donc exhaustive, avec ceux obtenus par un étiqueteur. En effet, un étiqueteur ne détectera jamais toutes les occurrences nominales, soit qu’il ne connaisse pas certains lexèmes (handicap qui s’aggrave encore pour le 18^e siècle), soit qu’il se trompe sur leur appartenance catégorielle, ce qui, en présence d’un adverbe de degré, n’est pas à exclure. Cependant, le pourcentage d’items intrinsèquement nominaux “ratés” par l’étiqueteur n’est pas si haut que cela, comme le montrent les quelques sondages que nous avons effectués. Ainsi, même pour la période 1700–1829, période à laquelle il n’a pas été entraîné, Tree Tagger identifie jusqu’à 78,57 % des types intrinsèquement nominaux après *trop* (14/17), soit 77,27% (17/21) des occurrences nominales. En outre, l’étiqueteur de Frantext ne donne presque aucun nom lorsqu’on lui soumet la requête *très* + adjectif (limitée à l’intervalle 1900–1913). C’est ce que nous apprend un contrôle manuel des 8915 occurrences de l’output, classées au préalable par le concordancier intégré à Sketch Engine, qui a permis de générer une liste de fréquence des items figurant à droite de l’adverbe.

est bien obligé de se fier aux choix faits par les concepteurs du corpus pour ce qui est de la représentation plus ou moins égale des différents genres à travers les différentes périodes. Par ailleurs, pour l’instant, l’interface de Frantext ne permet pas de vérifier la composition du corpus à partir de tableaux générés par genre et par période.

Seules 0,04 % des occurrences (soit 4/8915), soit 3 types (lexèmes) sur 1505, sont de nature nominale. On peut en conclure que presque tous les cas qui nous intéressent sont étiquetés par Frantext comme des noms, ce qui est rassurant. On peut en conclure que l'impact des étiqueteurs a dû être, malgré tout, réduit, et cela en dépit du contexte syntaxique (un adverbe de degré). Une deuxième asymétrie réside dans l'emploi de deux étiqueteurs différents, à savoir Tree Tagger pour 1700–1829 et l'étiqueteur de Frantext. Toutefois, un autre sondage montre que Tree Tagger parvient à identifier correctement 87,1 % (54 sur 62) des occurrences nominales détectées par Frantext pour la période 1885–1939, soit 84,1% des types (37 sur 44).⁵ En somme, ces quelques contrôles ont permis de conclure que les asymétries ne sont pas à ce point importantes qu'elles aient influencé outre mesure les résultats. Il n'empêche que nous nous abstenons de certains types de comparaisons.

Notons encore que le corpus Frantext catégorisé (après 1830) a été divisé en trois périodes. Cette opération nous permettait non seulement de suivre de plus près les développements entre 1830 et 1997, mais elle donnait en même temps des sous-corpus de taille comparable à celle du corpus 1700–1829, condition *sine qua non* pour l'application de certaines mesures de productivité. Pour répondre à cet impératif, il a fallu accepter que les tranches temporelles n'aient pas toutes la même étendue temporelle. Voici un aperçu de la taille des corpus:

Tableau 1: *Taille des corpus*

période	Corpus	Total occurrences
1560–1619	Frantext non catégorisé	7 482 630
1700–1829		47 602 286
1830–1884	Frantext catégorisé	40 942 615
1885–1939		44 508 708
1940–1997		41 539 427

Pour l'époque contemporaine, nous avons aussi effectué un sondage dans un corpus web (French Ten Ten corpus, Kilgarriff *et al.*, 2014) pour voir si les tendances de la dernière période de Frantext se confirmaient.

3. ORIGINE

3.1. *Les noms humains*, très gens de bien

Si la méthode que nous avons appliquée ne nous permet pas de remonter à la “première” attestation dans Frantext, elle nous permet néanmoins d'étayer un certain nombre d'hypothèses quant aux origines de la construction. Celle-ci semble trouver son origine dans le domaine des noms humains, et plus particulièrement des **noms humains évaluatifs**:

⁵ Nous n'avons pas pu faire l'expérience dans l'autre sens (lancer Frantext sur les noms relevés par Tree Tagger) pour la simple raison que l'étiqueteur de Frantext ne peut pas être utilisé indépendamment du corpus.

- (4) [...] prenez doresnavant en voz entreprisnes des conseillers **plus gens de bien** que ceux là. (Montaigne, 1592)

Ce domaine se présentait d'autant plus comme un terroir propice à l'écllosion de la construction dans la mesure où il était maximalelement exposé aux effets liés à certaines propriétés des stades plus anciens du français, que ce soit la plus grande liberté positionnelle, la non-obligatorité du déterminant ou la moindre rigidité de la catégorisation syntaxique (Carlier et Combettes, 2015).

Le premier argument qu'on peut avancer est d'ordre quantitatif:

Tableau 2: *Classes sémantiques*
(1560–1619)

classe N	Effectifs	%
N_humain	67	87,01
N_animal	1	1,30
N_inanimé	8	10,39
N_propre	1	1,30
Total	77	100,00

Les noms humains (et assimilés comme *diable*) étaient nettement dominants. Nous verrons que par la suite leur importance ne fera que diminuer. Voici les effectifs pour la période 1560–1619:

Gens_de_bien (20 occurrences), homme_de_bien (19), enfant (5), femme_de_bien (5), preud'homme (4), amant (3), diable (2), femme (2), gens_de_guerre (2), homme (2), capitaine_excellent (1), gentilhomme (1), vendeur (1)

Outre une très forte conventionnalisation – les 5 items les plus fréquents (tous humains) fournissent jusqu' à 68,83 % (53/77) des occurrences de la construction – on note la présence de nombreux termes évaluatifs, dont bon nombre sont des quasi-synonymes. Dans la plupart des cas, le sème évaluatif est réalisé de manière iconique sous la forme d'un composant caractérisant (un adjectif, un Sprép) à l'intérieur d'un syntagme souvent proche d'un nom composé⁶: *femme/homme/gens de bien, preud'homme, gentilhomme, capitaine excellent*.

Sur le plan sémantique, le sème évaluatif rapproche déjà ces N de la caractérisation, fonction prototypique de la catégorie adjectivale, dans la mesure où il met en place une échelle graduable. Sur le plan de la combinatoire syntaxique, cela a pu se traduire par une plus grande prédisposition à l'emploi avec un adverbe de degré, d'autant plus que plusieurs de ces noms humains évaluatifs comportaient en leur sein un formant caractérisant (*si preud' + homme, si gentil + homme*). En outre, rappelons que les adverbes de degré disposaient d'une liberté positionnelle

⁶ Nous avons cru utile de ne pas prendre en considération les occurrences de *gentil homme* dans lesquelles *gentil* était séparé (sans trait d'union) de *homme*, tout en sachant que le rendu des segmentations graphiques n'est pas toujours très stable dans les éditions de textes anciens.

plus grande qui leur permettait de s'éloigner du support modifié. En témoignent entre autres l'extraction tout à fait normale de l'adjectif en position préverbale (V2) d'une phrase attributive (Marchello-Nizia, 1996: 102)⁷ et ce qu'on appelle la 'quantification à distance' (Carlier, 2011) comme dans *j'ai assez vu de problèmes*. Il n'est, dès lors, pas interdit de penser que l'adverbe de degré pouvait également s'éloigner de l'élément caractérisant du N composé en se plaçant devant le nom au sein du (S)N, ce qui fait de ces N humains composés des candidats encore plus aptes à entrer dans la construction [Adv_{deg} + N]. En somme, l'ouverture de la construction [Adv_{deg} + Adj] à des éléments intrinsèquement nominaux doit être mise en lien avec d'autres évolutions structurelles du français, qui a vu augmenter le nombre de contraintes de contiguïté dans le domaine du SV et du SN et donc l'affirmation de la notion de syntagme hiérarchisé dans le cadre d'une langue plus configurationnelle (Carlier et Combettes, 2015). Il nous semble que l'émergence d'un syntagme adjectival plus cohésif pourrait également relever de cette tendance. Notons que ces noms humains qualitatifs à formant caractérisant constituent une classe bien particulière pour la syntaxe de l'époque, comme le montre aussi le fait qu'ils donnent lieu à des syntagmes très particuliers:

- (5) nous devons apprendre de c'est [*sic*] exemple, qu'il peut aduenir mesmes aux **plus gens de bien**. (de Combles, 1612)
- (6) Puisse la memoire de **si gens de bien** estre eternelle devant Dieu. (An., 1606)

Dans ceux-ci un nom humain évaluatif fonctionne comme syntaxiquement hybride, modifié par un adverbe de degré tout en étant précédé d'un déterminant. De ce fait, ce nom fonctionne comme tête adjectivale par rapport à ses dépendants, mais comme tête nominale par rapport à son intégration dans la phrase, puisque sur le plan externe, le syntagme se comporte comme un SN à part entière, qui peut saturer toutes sortes de positions nominales (et référentielles) comme sujet, COD, COI, complément du nom, etc. Ce fait a souvent été relevé⁸ (Haase, 1975 [1914]: 241; Gougenheim, 1951: 61; Moignet, 1974: 153, n. 23; Littré, 1872–1877; Grevisse et Goosse, 1991: 1463). On en trouve facilement des attestations dans Frantext et cela dès 1450 (*homme de bien, femme de bien, gens de bien*). Ce n'est d'ailleurs pas la seule surprise que nous réservent les adverbes de degré, qui, en plus, pouvaient être combinés entre eux:

⁷ Par exemple, *Mult par est proz danz Ogiers li Daneis* ('Très très est preux seigneur Ogier le Danois'; Marchello-Nizia, 1996: 98). Cette possibilité existait aussi pour la modification d'adjectifs intégrés au SN: *Car trop est mieudres chevaliers de vos*. ('Car il est bien meilleur chevalier que vous'; Marchello-Nizia, 1996: 100).

⁸ A ce propos, il semble que l'exemple emblématique de Tobler (1886: 121; repris par le *TLFi*), qui est censé témoigner de l'ancienneté de la combinaison adverbe de degré + N, doit être réexaminé: *Non ferés, dist Gaufrey, le tres mien ami chier* (De Gaufrey, mil. 13^e). Cet exemple n'est pas aussi convaincant qu'on le pense dans la mesure où il se pourrait très bien que *très* porte sur l'adjectif *chier*, ou encore, sur *mien* (B. Combettes, c.p.). L'origine préfixale de *très*, et la mobilité plus réduite que cela a pu entraîner, constitue un argument supplémentaire en faveur de *très mien*, d'autant plus que *mien* était une forme en voie de grammaticalisation qui se comportait encore comme un adjectif, comme le montre la coordination avec un modifieur adnominal (Carlier et Combettes, 2015: 53).

- (7) Quant à Guillaume son mary Il est **si treshomme de bien**, Qu'il ne se soucira de rien. (Jodelle, 1573)

On a là cependant un phénomène plus général qui s'observe aussi avec des têtes adjectivales et adverbiales (*si très fort*, *si très parfait*) et qui se rapprocherait de la composition (Gougenheim 1951: 57; Marchello-Nizia 1996), comme le montre aussi le fait que ces mots sont imprimés en un seul mot. En témoigne aussi le N composé assez bizarre *très-humilité*:

- (8) En vous re-supliant, en **très-humilité**, d'excuser [. . .]. (Papon, 1588)

Outre le poids numérique des noms humains et la présence d'un important sous-ensemble à sens qualitatif et autorisant des SN à modification adverbiale interne, on peut encore avancer un troisième argument qui fait des N humains une classe particulièrement disposée à entrer dans une construction typiquement adjectivale. En effet, comme ces noms qualitatifs relèvent de la classe sémantique des noms humains, dont on connaît la predisposition à la double appartenance catégorielle⁹ (adjectif et nom, cf. *supra*), ils sont massivement attestés avec les adverbes de degré. En voici le relevé exhaustif pour la période (1560–1619); notons qu'une part importante de ces noms contiennent un sème évaluatif:

amateur, ami, Aristotélien, artiste, Athée, badin, bragard, catholique, chrétien, coquin, débile, enchanteur, ennemi, esclave, expert, felon, gaillard, grue, guerrier, imbécile, libéral, maître (de), malotru, marâtre, marchand, ménager (de), menteuse, mignard, musard, mutin, paillard, phantasque, poltron, poupin, religieux, rêveur, saint, Sarrazin, savant, scelerat, seif, souverain, sujet (à), traître, trompeur, vagabond, vanteur, vierge, voisin, prisonnier

Ces noms bicatégoriels n'ont pas besoin du contexte [ADV_{degré} + X] pour fonctionner comme adjectif épithète, apposition ou attribut. C'est notamment leur aptitude à fonctionner même comme épithète qui les distingue des items [+ humains] intrinsèquement nominaux :¹⁰

- (9) et entr'eux un soldat **prisonnier** à cheval aussi (d'Audiguier, 1615)
(10) une plume **marastre** (Grévin, 1562)

Etant donné le statut bicatégoriel des ces N-Adj humains, les locuteurs/scripteurs de l'époque les rencontraient souvent en combinaison avec un adverbe de degré

⁹ Pour certains, cette double appartenance catégorielle remonte au latin (*amicus, vicinus*). Pour tous ces cas il est difficile de déterminer la catégorie de départ, comme le rappellent Carlier et Combettes (2015: 49, se référant e.a. à Ernout et Thomas, 1953: 165 sq.).

¹⁰ Etant donné les liens étroits entre le lexique bicatégoriel et les premiers items intrinsèquement nominaux recrutés par la construction [Adv_{deg} + N], on comprendra que la distinction entre bicatégoriel et monocatégoriel humain était parfois difficile à établir. C'est pourquoi il nous a paru prudent d'être assez sévère dans la délimitation de la classe des items intrinsèquement nominaux. En outre, nous nous sommes toujours basé sur les normes de l'époque, prenant comme point de départ les ouvrages lexicographiques de la période concernée.

en emploi adjectival. De là une analogisation¹¹ a pu se produire – basée sur le trait [+ humain] – qui a attiré de plus en plus de Nhumains intrinsèquement nominaux vers la construction adverbiale. Ceux-ci ont pu apparaître occasionnellement dans cette construction, ce qui a progressivement abouti à l'extension de la construction vers le domaine nominal (cf. § 4.).

En plus de leur similarité sémantique, les N intrinsèquement humains partageaient avec les N humains bicatégoriels la possibilité d'être construits sans déterminant en position d'attribut:

- (11) Charles fils de Louys Le Begue, estant encores **enfant**; n'avoit aage suffisant pour gouverner un si grand pays. (Fauchet, 1602)
- (12) Si nous sommes **gens de bien**, si nous aimons l'Etat (comme nous devons), [. . .]. (de Montchrestien, 1615)

Cet emploi était d'autant plus naturel qu'à cette époque l'emploi nu du N était encore plus libéral, notamment en position prédicative (cf. Gougenheim, 1951: 66; Brunot et Bruneau, 1949: 224). Rappelons d'ailleurs que même en emploi référentiel (argumental), la tendance à l'emploi de l'article commence à devenir systématique seulement à partir du 15^e siècle (Carlier et Combettes, 2015: 45). La construction attributive à nom nu non seulement rapproche les N intrinsèquement humains des N bicatégoriels, mais elle contribue, de manière plus générale, au brouillage des catégories dans l'esprit des locuteurs/scripteurs. En effet, dans cette position, l'opposition catégorielle est invisible. Le français en a gardé des traces, qui sont bien connues, par ailleurs:

- (13) Ils sont **amis** de longue date.

On comprend que la tentation de prêter un comportement complètement adjectival à ces N humains intrinsèquement nominaux en les faisant précéder d'un adverbe de degré en ait été d'autant plus grande. Nous verrons d'ailleurs que la disponibilité de la construction attributive nue semble également avoir joué un rôle dans l'emploi de N inanimés dans la construction [Adv_{deg} + N]. On constate, en outre, qu'elle a facilité encore par un autre biais l'émergence du patron [Adv_{deg} + N]. Si l'adverbe *plus* domine de loin les autres, c'est qu'il a pu se greffer sur une structure attributive à nom nu, quelle que soit d'ailleurs la classe du N tête, humaine ou inanimée (cf. Lauwers, 2006):

- (14) ce que tu m'as conté est **plus signe de haine** que de bon vouloir (J.de Lavardin, 1578)

La prédication attributive y sert de construction-hôte à la gradation avec l'adverbe *plus* (*que*). Il s'ensuit que la gradation n'affecte pas directement le nom (degré de N-ité) mais plutôt le degré d'adéquation de la prédication attributive, à tel point qu'on pourrait presque parler d'emploi métalinguistique ou énonciatif (en l'occurrence, dans quelle mesure peut-on parler de *haine*?). Dans cet

¹¹ A la suite de Traugott et Trousdale (2010: 38), nous distinguons ici l'*analogie* de l'*analogisation*, l'analogie étant le mécanisme cognitif ("analogical thinking") qui sous-tend le changement linguistique de ce nom, lequel est désormais appelé *analogisation*.

emploi – proche de *plutôt* (Grevisse et Goosse, 1986: 1463) – *plus* compare deux “qualités” (ou mieux, catégories) dans leur application à un seul référent:

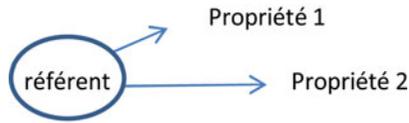


Figure 1: (Colour online) Un référent, deux échelles

Que cet emploi n’aille pas nécessairement de pair avec une adjectivation contextuelle ressort de la paire suivante:

- (15) Je suis **plus grimpeur** que rouleur. (Google, 31/08/2017)
 (15’) Christophe Moreau, qui est désormais **plus un grimpeur** qu’un rouleur. (Google, 31/08/2017)

En effet, l’emploi de *plus* est relativement indépendant du statut catégoriel du nom. Il apparaît donc que cet emploi de *plus* constitue une voie très facile pour faire accéder des noms à la construction [Adv_{deg} + N], qui a pu servir de transition vers des emplois qui affectent davantage¹² la nature catégorielle du nom nu.

3.2. Les autres classes de noms

Outre les N humains, qui semblent constituer le fonds héréditaire de la construction [Adv_{deg} + N], d’autres classes sémantiques sont attestées dès 1560–1619. Ainsi, on trouve un exemple d’un **N animal**:

- (16) ô **plus tygre** Belcar [= prince] que les tygres plus fiers (de Schelandre; 1608)

Il s’agit d’un emploi très emphatique qui reprend de manière presque tautologique la qualité graduée (N-ité) sous la forme d’un N étalon référent au prototype meilleur exemplaire (plus *fier* que les tygres les plus *fiers*, c’est-à-dire ‘cruels’¹³). Cet emploi repose sur une métaphore animale qui établit une ressemblance entre un homme et un animal, fondée sur le trait de la cruauté. C’est également une métaphore qui se trouve à la base de l’expression figée (*pas si*) *grue* (*que . . .*) qui est attestée 4 fois dans le corpus, ainsi que dans le DMF (en ligne). *Grue* a été métaphorisé dans son emploi nominal (animal > personne qui se laisse facilement tromper, femme facile (DMF)) et est devenu par la suite un adjectif à part entière, tout comme *dinde*, *vache*, *bête*, *bourrin*, . . . Enfin, la suite *trop*

¹²A l’instar de Grevisse et Goosse (1986: 1463), qui parlent d’un autre “phénomène”, on pourrait même se demander si on a affaire ici à la même construction de degré. Or, comme certains emplois d’autres adverbes pourraient également être considérés comme des emplois assez spécifiques (p.ex. *assez . . . pour: je ne suis pas assez N pour . . .*), nous avons préféré inclure ces exemples dans nos statistiques. Notons que le poids relatif de *plus* se réduit au fil du temps en faveur des autres adverbes: on passe de 51/77 (66,2%) pour la période 1560–1619 à 459/988 (46,4 %) pour la période 1830–1997.

¹³Je remercie le lecteur anonyme de *JFLS* qui a attiré mon attention sur la polysémie de *fier*.

animal a été analysée comme un emploi qualificatif d'un adjectif relationnel (donc bicatégoriel). Les animaux continueront à être utilisés dans la construction [Adv_{deg} + N] avec une fréquence plus ou moins stable. En voici les attestations pour la période 1700–1829 (9 ex.): *chien* (1704), *loup-garou* (1719), *tigre* (1728, 1757), *boeuf* (1755), *corbeau/chat-huant* (1756), *mouton* (1822).

Dans les **noms inanimés** (totalisant 8 occurrences), de leur côté, se reflètent les tendances que nous avons relevées pour les N humains. On trouve ainsi des N évaluatifs (pour le 18^e siècle encore *pas si bagatelles que . . .*):

- (17) ils l'expérimentent Père en **si grand'chose** (Calvin, 1560)
(18) disant ainsi parlant de leurs maistresses, ma divine, ma loyauté, ma beauté, [. . .], mais cella estoit **si gallimatias** et hors de mesure que rien plus (de Deimier, 1610),

ainsi que des N inanimés autorisant un emploi attributif nu, qui peuvent dès lors se combiner avec *plus*, dans le sens de 'plutôt' (cf. déjà l'ex. 14):

- (19) ce sont **plus herbes de vignes et de chemins**, ainsi que le lizet et la maurelle. (C. Estienne, 1564)

Que cette syntaxe "nue" fût naturelle à l'époque mais qu'elle le soit devenue moins par la suite, ressort clairement d'une correction qui apparaît à partir de l'édition de 1702 (Google Books): *des herbes de vignes et de chemins*.

Notons que l'adverbe *plus* apparaît aussi dans des emplois 'sérieux', voire métaphysiques où l'on juge du degré d'adéquation de l'application d'un concept (donc d'une catégorie) à un (type de) référent dans le cadre d'une comparaison avec un autre référent :¹⁴

- (20) Par exemple l'homme n'est pas **plus substance** qu'un grain de millet. (Dupleix, 1607)
(21) et ceste-cy [= la seconde espèce d'amitié] est **plus amitié** que la precedente, car il y a moins de disparité. (Charron, 1601)

Ainsi, deux référents sont comparés en termes de N-ité: dans quelle mesure réalisent-ils la catégorie désignée par le nom modifié par *plus*? On situe dès lors deux référents sur la même échelle:

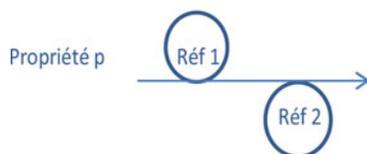


Figure 2: (Colour online) Deux référents, une échelle

¹⁴La présence assez remarquable de ce genre d'emplois pourrait s'expliquer par une surreprésentation de textes philosophiques dans le corpus pour cette période.

Une variante de ce type d'emploi se trouve dans les emplois tautologiques (impliquant aussi d'autres adverbes que *plus*) qui indiquent dans quelle mesure le référent réalise les traits de sa propre catégorie, comme dans cet exemple du début du 18^e siècle:

(22) Faisons une fiction **tres fiction**. (Dacier, 1714).

Une combinaison de ces deux types d'emplois – comparaison du degré de réalisation dans deux référents et catégorisation 'en boucle' s'appuyant sur un énoncé tautologique – se trouve en (16) et en (23), le seul exemple attesté d'un nom propre:

(23) cest acte [décapitation] le [Charles, duc d'Anjou) rendoit **plus Neron** que Neron, et plus Sarrazin que les Sarrazins. (de L'estoile, 1585–1587)

L'établissement du degré de réalisation de la catégorie se fait ici par le biais d'une comparaison avec le prototype meilleur exemplaire de la catégorie (en l'occurrence le tigre le plus fier et Néron en personne). Dans ces deux derniers exemples, on notera l'importance de la métaphore, qui les éloigne des emplois métaphysiques. Nous reviendrons sur cet aspect en 4.2.

3.3. Bilan intermédiaire

En somme, même si nous n'avons pas poussé nos recherches au-delà de 1560, nous avons pu formuler quelques hypothèses convergentes concernant le contexte d'apparition de la construction [Adv_{deg} + N]. Les noms humains, et notamment les noms humains évaluatifs comportant un élément caractérisant, semblent avoir été les meilleurs candidats à ouvrir la construction [Adv_{deg} + Adj] à des éléments nominaux. En plus de leur dominance numérique, on a noté leur proximité sémantique et en partie constructionnelle – la construction à attribut nominal nu, qui contribue d'ailleurs aussi au brouillage catégoriel – avec les lexèmes bicatégoriels, la présence du sème évaluatif, plus apte à la gradation et entraînant pour un sous-ensemble de ces N humains évaluatifs, ceux à composant caractérisant 'visible', des emplois syntaxiquement hybrides. En outre, on a relevé un emploi de *plus* qui n'affecte qu'indirectement le statut catégoriel du nom et qui constitue en quelque sorte un premier pas prudent vers des emplois pleinement recatégorisés. Les mêmes facteurs se retrouvent dans les quelques attestations de noms inanimés. De manière générale, nous avons noté la dominance d'exemples focalisés sur la catégorisation 'sérieuse' (parfois métaphysique), caractéristique qui sera reprise en 4.2. En filigrane, on a constaté aussi l'impact indirect de certaines propriétés structurelles plus générales de la langue de l'époque, notamment la liberté positionnelle de l'adverbe et l'absence encore plus généralisée de l'article.

4. DÉVELOPPEMENT DE LA CONSTRUCTION

Notons d'emblée qu'il y a une question à laquelle les asymétries relevées sous 2. nous empêchent de répondre: la construction en tant que telle devient-elle plus

fréquente, signe d'un ancrage cognitif plus fort? Nous verrons, toutefois, que là où les données sont comparables, on observe plutôt un recul en termes de fréquence absolue.

4.1. Les classes sémantiques

L'un des constats les plus frappants est la diversification des classes sémantiques du N apport:

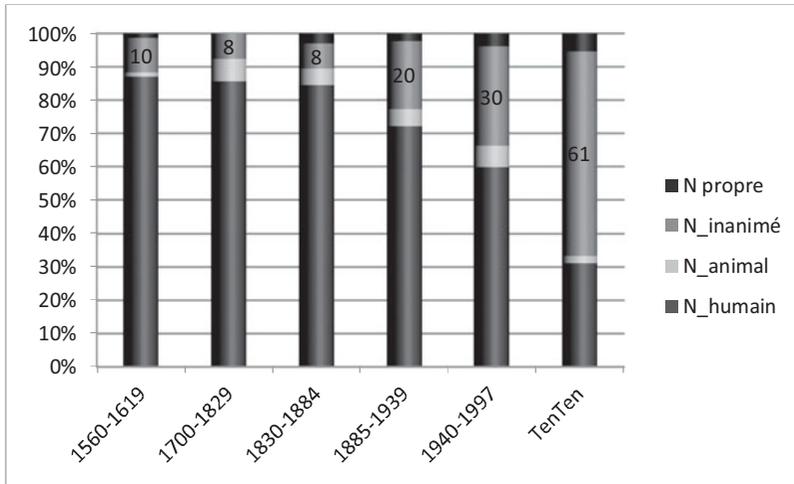


Figure 3: Poids relatif des classes sémantiques à travers le temps (occurrences)

De manière globale, on observe que la part des occurrences (angl. *tokens*) de noms inanimés s'accroît, tout comme celle des noms propres, au détriment des noms humains; la part des noms d'animaux reste plus ou moins stable. La courbe montante n'est cependant pas parfaite dans la mesure où on note même un petit recul des inanimés au début. Or, comme les nombres sont très bas (8 occurrences d'inanimés pour 1560–1619), il faudrait plutôt conclure à un statu quo jusqu'aux alentours de 1880. La même évolution, mais encore plus nette,¹⁵ apparaît au niveau de la représentation des classes sémantiques parmi les types lexicaux (i.e. le nombre de lexèmes différents): 32 %¹⁶ > 13 % > 17 % > 33 % > 41 %. Ce rééquilibrage

¹⁵Dans toutes les périodes, la représentation des N inanimés est plus forte au niveau des types qu'au niveau des occurrences, pour la simple raison que les N humains présentent plus de types à fréquence plus élevée, notamment dans les premières périodes (cf. infra). Ainsi, pour la période 1560–1619, le top 5 fournit jusqu'à 69,7 % des occurrences de la construction.

¹⁶Vu la haute fréquence des types humains dans la première période (voir note 10), les quelques items inanimés relevés reçoivent d'emblée un poids considérable par rapport au total des types (32 %). Notons que notre procédure de recherche a encore pu contribuer

des classes sémantiques à la suite de la montée en puissance des N inanimés se poursuit dans les registres moins contrôlés, comme le montre un sondage basé sur 100 occurrences randomisées du French Ten Ten corpus. Pas moins de 61% des occurrences sont désormais de type inanimé (voir la barre la plus à droite de la [figure 3](#)).

La catégorisation reportée sur la [figure 3](#) fait cependant encore abstraction de quelques nouvelles sous-classes qui ont émergé au fil du temps. En plus des noms humains collectifs tels que *gratin* ou *coeur de pension*, ce sont surtout les noms inanimés qui se diversifient au point d'inclure désormais aussi des séquences de mots syntaxiquement complexes qui réfèrent à des situations stéréotypées:

(24) une face [...] aux yeux très bons, mais, en revanche, très méchants, **très poing sur la hanche**, pour peu qu'un faquin les agace. (Verlaine, 1896)

Dans la même veine, l'ouverture vers les expressions délocutives (Benveniste, 1966: 277–285) dénotant une situation stéréotypique (marquant la perplexité) est un développement récent:

(25) et puis pour le fun, Code Geass gagne le prix de la scène **la plus WTF** [= what the fuck], School Days le prix de la fin la plus... Spéciale... (French Ten Ten Corpus)

Parfois les guillemets sont même conservés:

(26) J'étais [sic] **très 'fais comme tu veux, tout ce que tu veux, garde la [sic] comme maitresse, je m'en fous, mais reviens!!!!'** (Google, 31/08/2017)

Ce rééquilibrage au niveau des classes lexicales correspond aussi à une diversification des types, comme nous allons le voir en 4.3.

4.2. *L'interprétation associative (métaphorique)*

Les constats de nature quantitative ne sauraient toutefois occulter certaines évolutions qualitatives. Ainsi, dans la première période, la construction est dominée par des exemples comme

(27) lors qu'il estoit encore **si enfant** qu'il ne pouvoit avoir presque cognoissance du bien ny du mal. (d'Urfé, 1610)

Enfant y est pris au sens littéral du terme et le support est évalué en fonction de la catégorie 'enfant'. La phrase indique dans quelle mesure la catégorisation 'être (un) enfant' est justifiée. Il s'ensuit que le SN complet (*X est un enfant*) est impliqué

à une sous-évaluation du poids des inanimés pour les périodes après 1700, ce qui a pu creuser artificiellement encore l'écart avec la première période. En effet, si les N inanimés rencontrés dans la construction étaient plus variés que leurs homologues humains (plus standardisés), il se pourrait aussi que certains d'entre eux aient été plus rares et donc moins facilement repérables par des étiqueteurs (quoique ceux-ci aient également pu rater des N humains rares, bien sûr).

(test de Riegel, 1985). Jusqu'à 93,5 % (72 sur 77) des exemples de 1560 à 1619 sont de ce type.

Or, la même construction pouvait recevoir une deuxième interprétation, comme en (28), où quelqu'un demande au jeune Louis XIII lequel de deux adultes est encore le plus 'enfant':

- (28) M. de Montbazon et M. de Rosny y étoient. M. de Rosny lui demande: 'Monsieur, qui est **le plus enfant** de nous deux?' —C'est moucheu de Montbazon. (Héroard, 1605–1610)

Ici le SN complet n'est plus impliqué: M. de Montbazon n'est pas un enfant. Il s'agit ici d'une simple association métaphorique basée sur un certain degré de ressemblance. Cette interprétation, marginale au départ, est devenue dominante à l'heure actuelle, comme le montrent les chiffres de Frantext (1940–1997): 54,7 % (128/234).

Ce glissement vers le domaine purement associatif est très net dans les premiers exemples de N inanimés après 1830. Le ton a complètement changé, sans doute aussi sous l'influence du style romantique:

- (29) Il était là, grave, immobile, [...] C'était un regard fixe, et pourtant plein de trouble et de tumulte. Et à l'immobilité profonde de tout son corps, à peine agité par intervalles d'un frisson machinal, comme un arbre au vent, à la roideur de ses coudes **plus marbre** que la rampe où ils s'appuyaient, à voir le sourire pétrifié qui contractait son visage, on eût dit qu'il n'y avait plus dans Claude Frolo que les yeux du vivant. (Hugo, 1832)

Claude Frolo, l'archidiacre de Notre-Dame, observe Esméralda, la bohémienne, depuis son clocher. Pétrifié par ce spectacle, il fait pour ainsi dire corps avec la pierre du clocher: ses coudes sont plus marbre que la rampe. Le même mécanisme métaphorique est à l'œuvre dans le portrait de Mme de Staël:

- (30) mais elle l'avait [la parole, de Mme de Staël] plus vive encore et plus ardente. En un mot, elle était **plus tourbillon**. (Chenedolle, 1833)

La tendance à produire des associations plus libres, détachées de la catégorisation 'sérieuse', voire métaphysique, se manifeste aussi à travers l'émergence de noms propres. Hormis l'exemple isolé et assez particulier en (23), nous n'avons trouvé aucun N propre entre 1700 et 1829. Ensuite, la moisson s'enrichit. Rien que pour la période 1830–1851 on trouve: *Céladon* (un berger, personnage de l'Astrée), *Louis XIV* (2 occ.), *Louis XV*, *Régence*, [*prince de*] *Condé*, [*Journal des*] *Débats*. Les noms propres marquent un nouveau pas dans le renforcement de la tendance 'associative' dans la mesure où les associations qu'ils mettent en place reposent uniquement sur les connaissances encyclopédiques (des stéréotypes) que possédaient les locuteurs/scripteurs de l'époque à propos de tel référent illustre (personne, titre de journal, époque historique, ...). C'est le stéréotype social à l'état pur, pour ainsi dire, sans intervention du signifié (au sens structuraliste du terme). On est loin des énoncés linguistico-métaphysiques marquant l'adéquation du signifié contenu dans le N appoint.

4.3. La diversification au niveau des types lexicaux¹⁷

Que le rééquilibrage entre les classes sémantiques soit en réalité le résultat d'une diversification, ressort de manière encore plus nette de la multiplication des types lexicaux, pointant vers une productivité plus grande.

Le concept de productivité a d'abord été appliqué en morphologie (voir e.a. Baayen, 1993, et pour le domaine français, Dal, 2003), la productivité d'un suffixe (ou d'une règle de dérivation) étant le nombre de lexèmes qu'il/elle permet (et permettra) de générer. Ainsi, le suffixe *-ité* est plus productif que *-ise*. Ces dernières années, le concept a fait son entrée en syntaxe, notamment depuis l'avènement de la Grammaire des Constructions (e.a. Barðdal, 2008; Zeldes, 2012, Perek, 2016, . . .). La productivité d'une construction (syntaxique) consiste dans le nombre de lexèmes différents (i.e. les types) qui peuvent y entrer. Ainsi, on peut dire par exemple que la construction transitive (2156 verbes différents) est plus productive que la construction ditransitive en islandais, dont les effectifs se réduisent à 150 (Barðdal, 2008: 34).¹⁸ De ce fait, la productivité est encore plus que par le passé considérée comme un phénomène de nature gradable, qui peut être quantifiée.

De manière globale, on peut dire que la productivité de la construction [Adv_{deg} + N] a augmenté. Cela ressort tout d'abord du nombre de types relevés, qui monte de 1,26 (avant 1830) à 4 types par million de mots (après 1830):

Tableau 3: *Nombre des types* (type frequency)

	1700–1829	1830–1884	1885–1939	1940–1997
Taille corpus (mots)	47602286	40942615	44508708	41539427
n types dans Cx	60	167	187	165
Fq types N normalisée (par million de mots)	1,26	4,08	4,20	3,97

L'on notera que nous n'avons pas pu prendre en considération la première période, car le corpus n'a pas du tout la même taille (+/- 7,5 millions de mots). En effet, comme il est bien connu dans la littérature sur la productivité (Zeldes, 2012: 55), pour un phénomène donné, plus un corpus devient grand, plus il sera difficile de rencontrer de nouveaux types (pas encore recensés).¹⁹ C'est dire que le nombre de types est en partie fonction de la taille de l'échantillon. C'est ce qui est montré par les "vocabulary growth curves" (cf. Zeldes, 2012: 68) dans lesquelles on compte le

¹⁷Notons que nous avons dédoublé dans les comptages les N composés, p.ex. *femme* vs *femme du monde*.

¹⁸La productivité est un concept qui est entouré d'un certain flou et qui fait l'objet de débats auxquels nous ne pouvons pas faire écho ici (voir e.a. Dal, 2003; Barðdal, 2008, chap. 2; Zeldes, 2012). C'est notamment la tension entre *productivité réalisée* (nombre de types) et *productivité potentielle*, c'est-à-dire l'extensibilité de la construction à des nouveaux items lexicaux (et pour laquelle la proportion d'hapax passe pour une meilleure approximation) qui divise les esprits.

¹⁹Cela vaut aussi pour la *type frequency* normalisée.

nombre de types à chaque nombre de tokens vus. Cela dit, la fréquence normalisée des types du corpus 1700–1829 reste malgré tout légèrement sous-estimée à cause de la taille légèrement plus grande de ce corpus.

La productivité croissante de la construction ressort encore avec plus de netteté du rapport entre le nombre de types et le nombre d'occurrences, le *type/token ratio* (TTR), qui mesure la variation interne au sein de l'échantillon des occurrences (ou *tokens*), et donc la facilité avec laquelle la construction s'y applique à des items lexicaux différents, donc son ouverture sur le plan lexical. Nous y ajoutons aussi le *hapax/token ratio* (HTR), c'est-à-dire le nombre d'hapax (types à fréquence 1) par rapport au nombre total des occurrences. Cette dernière mesure est censée quantifier la productivité "potentielle" (Baayen, 2009: 902), c'est-à-dire la capacité de la construction à s'appliquer à des types lexicaux nouveaux (à l'instar des néologismes en morphologie), qui n'ont pas encore été rencontrés. Une portion plus importante d'hapax parmi les occurrences relevées montre que la construction donne facilement lieu à des applications inédites, qui sont analysées *in situ* (angl. "parsed") et qui, dès lors, ne sont pas stockées en mémoire comme entrées lexicales. C'est donc un indice très important de la 'vivacité' (et donc de la productivité) de la construction en tant que construction. Le dénombrement des hapax est d'autant plus important que tout corpus, aussi grand qu'il soit, est limité. Ainsi, si l'on parle de "types qui n'ont pas encore été rencontrés", cela veut dire en réalité qu'ils n'ont pas encore été attestés *dans le corpus utilisé*. De ce fait, le *hapax/token ratio* (HTR) s'avère être une meilleure façon d'extrapoler le comportement de la construction au-delà des limites du corpus. Cette mesure approche donc de plus près le véritable potentiel de la construction, ce qui est d'autant plus intéressant pour l'étude d'une construction qui a justement pour vocation de *transgresser* la catégorisation intrinsèque des items lexicaux. Voici les TTR et HTR:

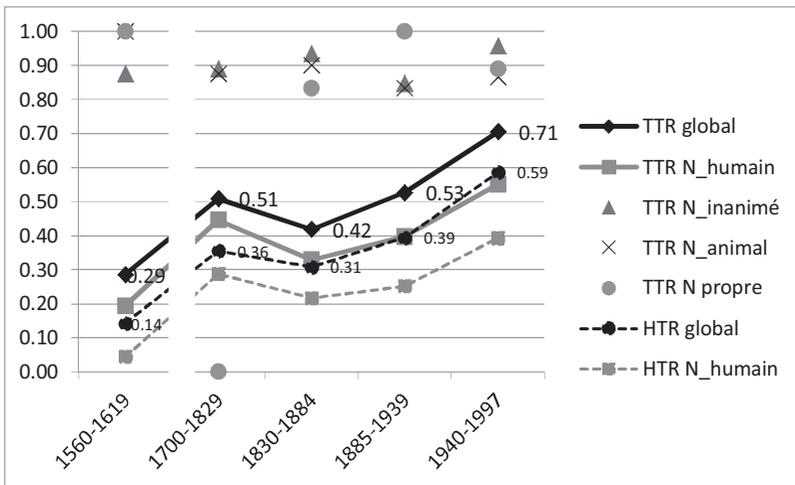


Figure 4: Type/token ratio (TTR) et hapax/token ratio (HTR)

On constate que, toutes classes sémantiques confondues, le TTR monte, tout comme le HTR (cf. TTR global et HTR global). Ainsi, dans la dernière période, jusqu'à 59 % des occurrences sont des types hapaxiques. Les noms humains suivent globalement la même évolution. Les types des autres classes sémantiques (représentant le plus souvent très peu d'occurrences, cf. 3.2.) ont toujours été très variés. En réalité, la montée est encore plus raide dans la mesure où le corpus 1560–1619 est de 5 à 6 fois plus petit, ce qui fait que les mesures TTR et HTR sont surévaluées pour cette période. En effet, dans un corpus plus grand, on pourrait s'attendre à retrouver de plus en plus de types déjà rencontrés, ce qui fait que le dénominateur de la fraction ne cesse de grandir, alors que numérateur "s'essoufle" pour ainsi dire.²⁰ La conclusion qui s'impose c'est que le lexique que l'on rencontre dans la construction s'est diversifié. On peut dès lors faire état de "host-class expansion" (Himmelman, 2004).

Il importe d'attirer l'attention sur le petit creux qui semble se dessiner entre 1830 et 1884. La construction aurait-elle plafonné à un moment donné avant de reprendre sa montée de plus belle? Il n'en est rien. On constate déjà que le recul du HTR (0.36 > 0.31) est moins prononcé que celui du TTR (0.51 > 0.42). Cela veut dire que la génération d'hapax se poursuit presque au même rythme. On en trouve la confirmation lorsqu'on compare le nombre d'hapax pour rapport au nombre de types:

Tableau 4: Hapax/type ratio

% hapax/types	1560–1619	1700–1829	1830–1884	1885–1939	1940–1997
toutes classes confondues	0,50	0,70	0,74	0,75	0,83
N humains	0,23	0,64	0,66	0,63	0,71

On constate même que la proportion des types hapaxiques parmi les types n'a jamais cessé d'augmenter. Comment expliquer ce paradoxe? En fait, le rapport hapax/type est moins sensible aux types à fréquence élevée – qui font gonfler le nombre d'occurrences (*tokens*) – que le rapport *type/token*, étant donné que la fréquence des types n'intervient pas dans le calcul. Il s'avère que c'est justement dans les types à fréquence élevée qu'un changement se produit. En effet, si l'on regarde de plus près le top 5 (ou le top 10) des types les plus fréquents, on constate que leur poids global se réduit au fil du temps:

²⁰Zeldes (2012: 65) illustre l'impact de la taille du corpus en montrant que le *hapax/token ratio* moyen pour le suffixe *-bar* en allemand dans 10 sous-corpus à taille égale est de 4%, alors que le *hapax/token ratio* de la construction pour tout le corpus n'est que de 0,9 %. En d'autres mots, un corpus 10 fois plus grand réduit le *hapax/token ratio* de plus de 4 fois.

Tableau 5: Token frequencies *des types les plus fréquents*

1700–1829			1830–1884			1885–1939			1940–1997		
	n	n/total		n	n/total		n	n/total		n	n/total
enfant	17	0,14	enfant	75	0,19	enfant	47	0,13	femme	11	0,05
diable	7	0,06	femme	27	0,07	femme	22	0,06	copain	10	0,04
femme	7	0,06	poète	21	0,05	homme	16	0,05	homme	9	0,04
homme_de_bien	7	0,06	homme	17	0,04	poète	10	0,03	enfant	6	0,03
héros	4	0,03	gentilhomme	14	0,04	peintre	7	0,02	mère	6	0,03
TOTAL top 5		0,36			0,39			0,29			0,18
homme	4	0,03	diable	9	0,02	théâtre	6	0,02	minot	5	0,02
roi	4	0,03	amant	7	0,02	gosse	5	0,01	filles	3	0,01
soldat	4	0,03	ange	7	0,02	Monsieur	5	0,01	homme_du_monde	3	0,01
gens_de_bien	3	0,03	homme_du_monde	6	0,02	bébé	4	0,01	journaliste	3	0,01
poète	3	0,03	mère	6	0,02	copain	4	0,01	poète	3	0,01
TOTAL top 10		0,51			0,47			0,35			0,25
n total d'occ.	118			399			355			234	

Ainsi, entre 1830 et 1884 le top 5 fournissait encore 39 % des occurrences (chiffre comparable à la période 1700–1829), alors que ce chiffre retombe à 18 % dans la période 1940–1997. On pourrait dire que le sommet de la pyramide s’est effrité, alors que sa base n’a cessé de s’élargir (cf. HTR et hapax/type ratio). L’impact de cette évolution est telle que la fréquence globale de la construction s’en est ressentie:

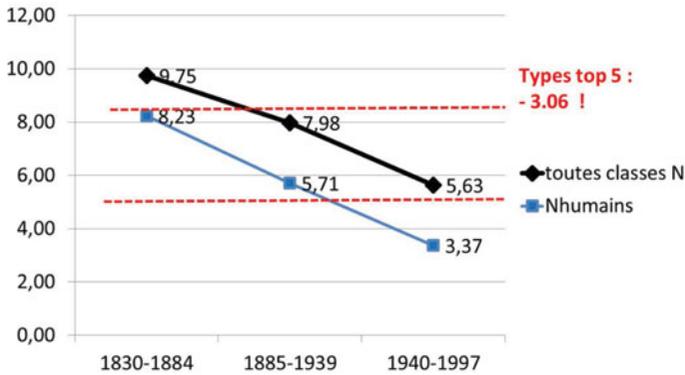


Figure 5: (Colour online) Token frequency de la construction [$Adv_{deg} + N$]

Si l’on considère l’évolution rien que pour les 5 items les plus fréquents de la période 1830–1884, on constate que ceux-ci reculent de 3,06 occurrences/million de mots, ce qui correspond à 63% de l’écart global entre 1830–1885 et 1940–1997. A regarder de plus près ces items, on constate qu’ils appartiennent tous au fonds héréditaire [+ humain] de la construction: *enfant, femme, poète, homme et gentilhomme*. Notamment la chute vertigineuse d’*enfant* est remarquable. En somme, alors que la construction se diversifiait progressivement (même au sein des N humains) et cela de plus en vite, elle a conservé le noyau de son fonds lexical héréditaire jusqu’en 1885, avant de le repousser progressivement, ou du moins, de le réduire à la portion congrue. D’où ce constat assez curieux selon lequel la construction devient moins fréquente, tout en étant de plus en plus diversifiée, et de plus en plus à même de générer des hapax, qui font un bond assez remarquable après 1940: 83 % des types sont désormais des hapax.

5. CONCLUSION

La construction [$Adv_{deg} + N$] semble être une extension du patron [$Adv_{deg} + Adj$] sous la pression d’un mouvement d’analogisation qui poussait les locuteurs/scripteurs à utiliser la classe des N humains monocatégoriels dans une construction dans laquelle ils rencontraient souvent les items intrinsèquement bicatégoriels de la classe. Parmi les premiers items à franchir la barre on retrouve notamment un important sous-groupe, des noms humains évaluatifs – déjà plus aptes

à la gradation – comportant dans leur forme un composant évaluatif (*homme de bien*), qui pouvaient même fonctionner comme tête syntaxiquement hybride d'un SN. Ces évolutions ont également été favorisées par certaines propriétés structurelles plus générales du français de l'époque, telles que la liberté positionnelle plus grande des adverbes de degré et l'absence de l'article. Par la suite, et notamment après 1885, cette construction s'est diversifiée, comme le montrent la représentation plus équilibrée des classes sémantiques (avec la montée des N inanimés) et la diversité grandissante du lexique nominal, et notamment la proportion croissante d'hapax. Cette expansion de la classe-hôte (*host-class expansion*, Himmelmann, 2004) est typique d'une construction qui gagne en productivité. Parallèlement, elle a jeté par dessus bord son fonds lexical héréditaire, à tel point que, dans l'ensemble, on la retrouve moins fréquemment dans le discours. Aussi paradoxal que cela puisse paraître, on a là la preuve que la construction gagne en maturité. Elle apparaît désormais comme moins conventionnalisée (lexicalisée) et gagne donc en schématicité (Langacker, 1987). Que le champ d'application de la construction soit devenu plus vaste apparaît aussi dans le fait qu'elle s'utilise de nos jours surtout pour marquer la ressemblance, dans le cadre d'associations de concepts assez libérales fondées sur des stéréotypes sociaux, qui en circonscrivent en même temps les limites pragmatiques. Dans le même sens, il faut interpréter l'apparition plus fréquente de noms propres (l'association encyclopédique à l'état pur, pour ainsi dire) et l'ouverture vers le domaine d'expressions délocutives mettant en jeu des analogies situationnelles. Ces évolutions ont fait de cette construction un puissant outil de recatégorisation contextuelle qui répond au besoin expressif de caractériser des référents de façon à la fois approximative et originale, sur la base d'un faisceau mal défini de propriétés descriptives stéréotypiques associées au concept nominal.

Adresse pour correspondance:
e-mail: peter.lauwers@ugent.be

RÉFÉRENCES

CORPUS, OUTILS INFORMATIQUES ET SITES WEB

AntConc: <http://www.laurenceanthony.net/software/antconc/>

DMF: *Dictionnaire du Moyen français*, <http://www.atilf.fr/dmf/>

Frantext: www.frantext.fr/

Huguet, E. 1925–1967. *Dictionnaire de la langue française du seizième siècle*. Paris: Champion.

Le Petit Robert: dictionnaire alphabétique et analogique de la langue française. 2010. Paris: Dictionnaires Le Robert.

Le Trésor de la Langue Française informatisé (= TLFi): <http://atilf.atilf.fr/>

Sketch, French Ten Ten Corpus: <https://www.sketchengine.co.uk/>

OUVRAGES ET ARTICLES

- Amiot, D. and Van Goethem, K. (2012). A constructional account of French *-clé* 'key' and Dutch *sleutel* 'key' in *mot-clé/sleutelwoord* 'key-word'. *Morphology*, 22/3: 347–364.
- Baayen, H. (1993). On Frequency, Transparency and Productivity. In: G. Booij and J. Van Marle (dir.), *Yearbook of Morphology 1992*. Dordrecht: Springer, pp. 181–208.
- Baayen, H. (2009). Corpus linguistics in Morphology: morphological productivity. In: A. Lüdeling and M. Kytö (dir.), *Corpus Linguistics. An International Handbook*. Berlin: De Gruyter, pp. 899–919.
- Barðdal, J. (2008). *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: Benjamins.
- Benveniste, É. (1966 [1958]). Les verbes délocutifs. In: *Problèmes de linguistique générale*, vol. I. Paris: Gallimard, pp. 277–285.
- Brunot, F. et Bruneau, Ch. (1949). *Précis de grammaire historique de la langue française*. Paris: Masson.
- Carlier, A. (2011). From *multum* to *beaucoup*: between adverb and nominal determiner. In: L. Toveana (dir.), *Determiners in and across Time*. Londres: College Publications, pp. 55–87.
- Carlier, A. et Combettes, B. (2015). Typologie et catégorisation morphosyntaxique: Du latin au français moderne. *Langue Française*, 187: 15–58.
- Dal, G. (2003). Productivité morphologique: définitions et notions connexes. *Langue Française*, 140: 3–23.
- Ernout, A. et Thomas, F. (1953²). *Syntaxe latine*. Paris: Klincksieck.
- Goes, J. (1999). *L'adjectif. Entre nom et verbe*. Paris/Bruxelles: Duculot.
- Gougenheim, G. (1951). *Grammaire de la langue française du seizième siècle*. Lyon: IAC.
- Grevisse, M. et Goosse, A. (1991¹²). *Le Bon Usage*. Paris: Duculot.
- Haase, A. (1975 [1914]). *Syntaxe française du XVIIe siècle*. (Nouv. éd. trad. et remaniée par M. Obert).
- Himmelmann, N. P. (2004). Lexicalization and grammaticization: Opposite or orthogonal? In: W. Bisang, N. P. Himmelmann et B. Wiemer (dir.), *What Makes Grammaticalization: A Look from its Components and its Fringes*. Berlin: Mouton de Gruyter, pp. 21–42.
- Kerleroux, F. (1996). *La coupure invisible. Etudes de syntaxe et de morphologie*. Paris: Presses Universitaires du Septentrion.
- Kerleroux, F. (1999). Identification d'un procédé morphologique: la conversion. *Faits de Langue*, 14: 89–100.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1: 7–36.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- Lauwers, P. (2006). Les emplois lexicalisés de l'article zéro en fonction d'attribut du sujet. *Cahiers de lexicologie*, 89/2: 29–46.
- Lauwers, P. et Willems, D. (2011). Coercion: Definition and challenges, current approaches, and new trends. *Linguistics*, 49/6: 1219–1235.
- Lauwers, P. (2014a). Between adjective and noun: category/function mismatch, constructional overrides and coercion. In: F. Masini et R. Simone (dir.), *Word classes*. Amsterdam: Benjamins, pp. 203–225.

- Lauwers, P. (2014b). Copular constructions and adjectival uses of bare nouns in French: a case of syntactic recategorization? *Word*, 60: 89–122. [Preprint in French: 2005]
- López Díaz, M. (2008). Noms communs et noms propres 'qualitatifs'? *Travaux de Linguistique*, 56: 69–95.
- Marchello-Nizia, C. (1996). La place de l'attribut du sujet et son évolution aux XII^e et XIII^e siècles en français. *Langue Française*, 111: 97–108.
- Moignet, G. (1974). *Études de psycho-systématique française*. Paris: Klincksieck.
- Noailly, M. (1990). *Le substantif épithète*. Paris: P.U.F.
- Perek, F. (2016). Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics*, 54(1): 149–188.
- Riegel, M. (1985). *L'adjectif attribut*. Paris: P.U.F.
- Tobler, A. (1886). *Vermischte Beiträge zur französischen Grammatik*, tome 1. Leipzig: Hirzel.
- Traugott, E. Closs et Trousdale, G. (2010). *Gradience, gradualness and grammaticalization: How do they intersect?* In: E. Closs Traugott et G. Trousdale (dir.), *Gradience, Gradualness and Grammaticalization*, Amsterdam: Benjamins, pp. 19–44.
- Zeldes, A. (2012). *Productivity in Argument Selection: From Morphology to Syntax*. Berlin: Mouton De Gruyter.