



La concurrence affixale en diachronie : le cas des préfixes de haut degré en français

Emmanuel Cartier, Richard Huyghe

► To cite this version:

Emmanuel Cartier, Richard Huyghe. La concurrence affixale en diachronie : le cas des préfixes de haut degré en français. *Linx*, 2021, <10.4000/linx.8078>. <hal-03598157>

HAL Id: hal-03598157

<https://hal.science/hal-03598157v1>

Submitted on 4 Mar 2022

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.



HAL Authorization



Linx

Revue des linguistes de l'université Paris X Nanterre

82 | 2021

Entre vieillissement et innovation : le changement linguistique

La concurrence affixale en diachronie : le cas des préfixes de haut degré en français

Affix rivalry in diachrony: a case study of French evaluative prefixes

Emmanuel Cartier et Richard Huyghe



Édition électronique

URL : <https://journals.openedition.org/linx/8078>

DOI : [10.4000/linx.8078](https://doi.org/10.4000/linx.8078)

ISSN : 2118-9692

Éditeur

Presses universitaires de Paris Nanterre

Référence électronique

Emmanuel Cartier et Richard Huyghe, « La concurrence affixale en diachronie : le cas des préfixes de haut degré en français », *Linx* [En ligne], 82 | 2021, mis en ligne le 15 juillet 2021, consulté le 20 juillet 2021. URL : <http://journals.openedition.org/linx/8078> ; DOI : <https://doi.org/10.4000/linx.8078>

Ce document a été généré automatiquement le 20 juillet 2021.

Département de Sciences du langage, Université Paris Ouest

La concurrence affixale en diachronie : le cas des préfixes de haut degré en français

Affix rivalry in diachrony: a case study of French evaluative prefixes

Emmanuel Cartier et Richard Huyghe

Introduction¹

- 1 Les préfixes comme *archi-*, *hyper-*, *méga-*, *super-*, *ultra-*, qui expriment le haut degré (Guilbert & Dubois, 1961 ; Amiot, 2004 ; Izert, 2011) abondent en français, et on peut s'interroger sur les raisons d'être de ce foisonnement formel. Dans une conception systémique de la langue, postulant l'existence de différences irréductibles entre les formes existantes, on s'attend à ce que ces différents préfixes se distinguent par des propriétés d'ordre structurel (différences de construction phonologique, syntaxique ou sémantique) ou variationnel (spécificités diaphasiques, diastratiques, diatopiques, etc.). Cependant, l'hypothèse de formes générées en surabondance, sans nécessaire distinctivité, doit également être examinée. Certains travaux récents suggèrent en effet que la dérivation (Fradin, 2016 ; Dal et al., 2018 ; Bonami & Strnadova, 2019), à l'instar de la flexion (Thornton, 2012 et 2019), n'exclut pas l'existence de formes redondantes dans les paradigmes morphologiques. Or les préfixes de haut degré (désormais PHD) pourraient constituer un cas de ce type.
- 2 Pour apporter quelques éléments de réflexion à ce débat, nous nous interrogeons ici sur l'évolution diachronique des PHD en français moderne. On est frappé en effet par la productivité actuelle de ces préfixes, certains d'entre eux se rencontrant fréquemment dans les formations néologiques, telles que celles sous (1).
(1) *ultra-rural*, *ultra-pollueur*, *méga-salaire*, *méga-motivé*, *extra-nourrissant*, *extra-riche*, *hyper-fierté*, *hyper-tendance*, *supra-pouvoirs*, *supra-culte*²
- 3 On peut se demander comment évolue globalement l'emploi des PHD, et si cette évolution est marquée par des variations importantes de la fréquence d'emploi des

différents PHD. Il s'agit de savoir, d'une manière générale, comment s'organise la diversité des PHD en diachronie, et si l'on constate des phénomènes d'émergence ou d'effacement de certains préfixes dans le temps.

- 4 Pour répondre à ces questions, nous observerons le comportement de huit PHD (*archi-*, *extra-*, *hyper-*, *maxi-*, *méga-*, *super-*, *supra-*, *ultra-*)³ sur la période 1800-2010, en nous basant des données recueillies au moyen de *Google Ngram*⁴. Notre étude procédera en trois temps. Nous reviendrons d'abord sur les propriétés générales de construction des PHD en français, et sur leur coexistence dans le système morphologique. Nous présenterons ensuite notre méthodologie d'extraction et de traitement des données. Enfin, nous exposerons les résultats de nos observations, et proposerons une analyse des évolutions d'emploi collectives, individuelles et comparées des huit préfixes étudiés.

1. Les préfixes de haut degré : concurrence et polyvalence

- 5 L'expression du haut degré en français passe par différents procédés lexicaux et grammaticaux. Comme l'indiquent entre autres Riegel *et al.* (1994), Amiot (2004) et Dostie (2018), le haut degré en français peut être exprimé par des affixes (*hyper-*, *-issime*), des adverbes (*extrêmement*, *infiniment*), des adjectifs (*énorme*, *colossal*), des locutions (*au bout*, *à l'os*), ainsi que par la reduplication (*fou fou fou*), l'exclamation (*Que c'est beau!*) ou la prosodie (accent d'insistance). Dans le cas des PHD, les formes sont abondantes et paraissent souvent interchangeables, comme dans les exemples (2).
 - (2) a. Pierre est {*méga-* / *hyper-* / *archi-* / *super-* / *extra-* / *ultra-*}sympa.
 - b. C'est un {*méga-* / *hyper-* / *archi-* / *super-* / *supra-* / *ultra-*}mensonge.⁵
- 6 On peut estimer que *super-* exprime un moindre degré que les autres PHD, mais l'hypothèse d'une différenciation claire des PHD selon le degré d'intensité paraît hasardeuse. L'interprétation des variantes sous (2) paraît très proche, *archi-*, *extra-*, *hyper-*, *méga-*, *ultra-*, *supra-* ayant pour fonction principale ici de décrire le degré très élevé d'une propriété donnée.
- 7 La question demeure de savoir si les différents PHD ne se distinguent pas par leur construction morphologique (s'agissant notamment de la sélection des bases de dérivation) ou par leurs capacités d'emploi respectives, que ce soit pour exprimer d'autres propriétés que le haut degré, ou pour se lexicaliser dans des emplois non affixaux.

1.1. Construction morphosémantique

- 8 Les PHD ont des origines étymologiques semblables. Comme l'indique Amiot (2004), ils sont issus principalement de prépositions locatives latines (*extra-*, *hyper-*, *super-*, *supra-*, *ultra-*), voire d'adjectifs du grec ou du latin (*maxi-*, *méga-*). Leurs conditions de construction morphologique sont également comparables, pour ce qui est de la sélection grammaticale des bases de dérivation, celles-ci étant très majoritairement des adjectifs et des noms (cf. tableau 1). La construction marginale avec des adverbes de haute fréquence est observée pour la plupart des PHD (à l'exception notable de *maxi-*), et on recense quelques formations sur base verbale. Il paraît toutefois difficile, dans ces

deux derniers cas, d'établir une régularité constructionnelle, notamment en raison de la variation graphique et de l'analysabilité des PHD comme adverbes (cf. *infra*).

Tab. 1 : Catégorie grammaticale des bases de dérivation sélectionnées par les PHD

Préfixe	Base Adj	Base N	Base Autre
ARCHI-	archifacile	archi-domination	archi-contre
EXTRA-	extrafin	extra-fraîcheur	extra-loin
HYPER-	hyperdoué	hyperstress	hyper-ventiler
MAXI-	maxi-long	maxi-frayeur	-
MÉGA-	méga-cool	méga-pollution	méga-vite
SUPER-	supercritique	super-tempête	supermécaniser
SUPRA-	supra-intelligent	supravigilance	supra-mal
ULTRA-	ultrascret	ultra-champion	ultra-tard

- 9 La construction sémantique des dérivés varie selon la catégorie de la base. Lorsque la base est un adjectif, la prédication de haut degré associée à l'affixe s'établit directement à partir de la gradabilité de la propriété décrite par l'adjectif (ex. *extradoux*, *ultraréaliste*, *hypertendu*). Lorsque la base est un nom, deux cas de figure se présentent. Si le nom est lui-même intensif, i.e. comprenant dans sa structure sémantique une variable d'intensité (Flaux & Van de Velde, 2000), alors la prédication de haut degré porte sur la variable en question 50 c'est typiquement le cas avec des noms de propriété, dont beaucoup sont d'ailleurs dérivés d'adjectifs (ex. *hypersensibilité*, *superfluidité*). Si au contraire le nom de base n'est pas un nom intensif, alors la prédication de haut degré s'établit en fonction d'une propriété gradable prototypiquement associée à l'entité décrite (ex. l'héroïsme dans *superhéros*, la tromperie dans *archi-mensonge*, l'isolement dans *ultra-niche*, la théâtralité dans *hyper-spectacle*). Dans les cas de formation sur base verbale ou adverbale, la construction sémantique opère généralement sur le même modèle qu'avec les adjectifs, i.e. par la modification d'une composante intensive incluse dans le sémantisme de la base (ex. la rapidité dans *méga-vite*, le degré d'oxygénation dans *hyper-ventiler*). D'après nos observations, il n'y a pas de différence majeure entre les différents PHD dans la sélection grammaticale des bases ni dans l'élaboration sémantique de dérivés exprimant le haut degré. L'absence de telles spécificités, ainsi que de différences morphophonologiques apparentes, laisse penser qu'on a affaire à des formes très proches, sinon équivalentes.

1.2. Polysémie

- 10 Les principales propriétés distinctives des différents PHD sont à rechercher du côté des emplois qui ne relèvent pas, à proprement parler, de l'expression du haut degré. Les PHD peuvent en effet intervenir dans la formation de lexèmes pour indiquer d'autres modifications sémantiques que l'intensification. Par exemple, *archi-* peut indiquer une position supérieure dans une hiérarchie (*archiduc*), *extra-* peut exprimer l'extériorité (*extraterritorial*), *méga-* peut spécifier la grandeur d'une unité de mesure (*mégawatt*), etc. Les types sémantiques des bases de dérivation varient d'après ces constructions particulières (ex. *archi-* pourra sélectionner des noms de position hiérarchique, *extra-* des adjectifs relationnels, *méga-* des noms d'unité de mesure, etc.). Amiot (2004) propose un récapitulatif de cette diversité sémantique dans le tableau reproduit ci-dessous (tableau 2).

Tab. 2 : Polyvalence d'emploi des préfixes de haut degré (d'après Amiot 2004)

	loc. spatiale	loc. non spatiale	sup. hiérar	supériorité	excès	haut degré
<i>archi-</i>			<i>archidiacre</i> <i>archiduc</i>			<i>archi-nul</i> <i>archi-sévère</i>
<i>extra-</i>	<i>extra-territorial</i> <i>extra-corporel</i>	<i>extra-universitaire</i> <i>extra-conjugal</i>				<i>extra-fin</i> <i>extra-souple</i>
<i>hyper-</i>				<i>hypermolécule</i> <i>hyperespace</i>	<i>hypertension</i> <i>hyperchloxydrie</i>	<i>hypernerveux</i> <i>hyperraffiné</i>
<i>super-</i>	<i>superstructure</i>	<i>supersonique</i> <i>super-léger</i>		<i>superbombardier</i> <i>superovulation</i>		<i>superlong</i> <i>superfin</i>
<i>sur-</i>	<i>surveste</i> <i>surnappe</i>		<i>surintendant</i> <i>surarbitre</i>		<i>suralimentation</i> <i>surévaluer</i>	<i>surexcité</i> <i>suraigu</i>
<i>ultra-</i>		<i>ultraviolet</i> <i>ultrason</i>				<i>ultra-libéral</i> <i>ultra-court</i>

- 11 Les différentes significations distinguées ont toutes un lien conceptuel avec le haut degré, et les opérations de dérivation mettant en jeu les PHD peuvent en ce sens être considérées comme polysémiques. On peut faire l'hypothèse que dans un certain nombre de cas, le sens de haut degré est élaboré par métaphore à partir d'un sens locatif. La description d'ensemble proposée par Amiot demanderait un examen minutieux, et certaines distinctions ou certains exemples donnés pourraient être discutés. Par exemple, la frontière entre le sens d'excès, voire de supériorité, et celui de haut degré paraît difficile à tracer ²⁰ en l'occurrence, la distinction esquissée par Amiot semble dépendre de la catégorie grammaticale de la base et du dérivé (excès et supériorité pour les noms vs haut degré pour les adjectifs). Il n'en reste pas moins que la diversité d'emploi et l'existence de combinaisons d'interprétations propres à chaque PHD sont manifestes. On peut penser que la persistance des différents PHD dans le système dérivationnel tient notamment à ces variations de structure polysémique, et que les emplois de haut degré constituent potentiellement des cas de neutralisation des différences entre PHD.

1.3. Lexicalisation

- 12 Outre les différences de configuration polysémique, les PHD se distinguent les uns des autres par des capacités de lexicalisation variées ²¹ si par *lexicalisation*, on entend le passage d'un élément grammatical à un élément lexical (van der Auwera, 2002). L'emploi des PHD comme adverbes de haut degré paraît certes généralisable, mais il est difficile à distinguer de l'emploi affixal, comme le signalent Guilbert et Dubois (1961). Dans une phrase comme (3), le statut préfixal ou adverbial de *super*, *hyper*, *méga* et *ultra* est discutable, et la réanalyse grammaticale est dans tous les cas possible.

(3) Elle est {super / hyper / méga / ultra} (-) intelligente.

- 13 Le traitement graphique, dans l'hypothèse où l'espace correspondrait à l'emploi adverbial, et la fusion ou le trait d'union à l'emploi affixal, est un indice difficilement exploitable, tant il paraît aléatoire et non systématisé à l'écrit. Par exemple, la scission graphique se rencontre avec les noms dans des cas où il n'y a pas de lexicalisation adjectivale possible des PHD, comme dans (4a), où *hyper* ne peut pas être analysé comme un adjectif épithète (4b).

(4) a. Dans la même mouvance, ils veulent de l'hyper choix, tout en entrant dans l'ère de l'hyper tranquillité: pas de pression des vendeurs, liberté de choisir. (frWaC)

b. *un choix très hyper, *une tranquillité vraiment hyper

- 14 Contrairement à l'emploi adverbial, l'emploi comme adjectif de haut degré est non généralisable, et réservé à certains PHD (*super* et *extra*) :
- (5) Le résultat est {*super* / *extra* / ??*archi* / ??*hyper* / ??*maxi* / ??*méga* / ??*supra* / ??*ultra*}.
- 15 Il en va de même pour la lexicalisation comme nom, qui résulte généralement d'une opération de conversion et/ou de troncation (ex. l'adjectif *extra* se convertit en nom, *maximum* s'apocope en *maxi*) :
- (6) a. Ils font un *extra* en allant au restaurant.
 b. On a atteint le *maxi* de la puissance autorisée.
 c. Pierre a fait le plein de *super*.
 d. Les *ultras* ont tenté de pénétrer dans le bâtiment.
- 16 Indépendamment de ces particularités lexicales, les PHD se concurrencent directement et sans distinction claire dans l'expression du haut degré. La concurrence dérivationnelle a fait l'objet de nombreux travaux dans la période récente (cf. Fábregas, 2010 ; Lindsay & Aronoff, 2013 ; Arndt Lappe, 2014 ; Schulte, 2015 ; Aronoff, 2016 ; Varvara, 2018 ; Dressler *et al.*, 2019 ; Gardani *et al.*, 2019 ; Naccarato, 2019 ; *inter alia*). Les études sur le français portent principalement sur la suffixation, et la perspective adoptée est le plus souvent synchronique (cf. Ferret *et al.*, 2010 ; Martin, 2010 ; Koehl & Lignon, 2014 ; Fradin, 2016 ; Dal *et al.*, 2018 ; Bonami & Thuilier, 2019 ; Missud & Villoing, 2020 ; Huyghe & Wauquier, 2021 ; *inter alia*). Un des principaux objectifs de ces études est de mettre en évidence les propriétés distinctives des affixes concurrents, mais l'hypothèse de la surabondance affixale est dans certains cas examinée. On peut noter à cet égard que l'existence en synchronie de formes surabondantes ne contredit pas l'hypothèse d'une nécessaire répartition complémentaire des paires forme-sens, si la situation de surabondance constitue une phase transitoire entre deux états de langue dans lesquels la répartition est effective. Aussi la surabondance systémique ne peut-elle être observée qu'en diachronie. Dans le cadre de l'étude des PHD, nous souhaitons examiner l'évolution dans le temps de la productivité relative des différents affixes, et l'existence ou non de transferts de distribution entre les formes concurrentes.

2. Méthodologie

- 17 Dans cette section, nous présentons le corpus sur lequel se base notre étude, le prétraitement effectué afin de le rendre exploitable, quelques informations générales concernant les données issues de ce prétraitement, ainsi que les outils méthodologiques utilisés dans l'analyse diachronique des PHD.

2.1. Corpus et prétraitement⁶

- 18 Pour étudier les PHD, nous exploitons l'outil *Google Ngram* (désormais GNG) (Michel *et al.*, 2010 ; Lin *et al.*, 2012), donnant librement accès aux *n-grams*⁷ (de 1 à 5) ayant sur l'ensemble de la période 1600-2010 une fréquence supérieure à 40, et extraits de la plus grande collection disponible de documents numérisés (corpus *Google Books*). Les documents concernés sont principalement des ouvrages littéraires et des documents scientifiques et techniques. Pour chaque *n-gram*, la forme brute, la fréquence absolue, le nombre de documents concernés et l'année sont renseignés. Nous exploitons les données disponibles pour le français moderne et contemporain, dans la période allant

de 1800 à 2010 (soit plus de 100 milliards d'occurrences de mots). La procédure suivie pour extraire de ce corpus les constructions incluant l'un des huit PHD retenus comprend quatre étapes principales.

- 19 **Étape 1.** Nous récupérons et filtrons automatiquement les données brutes du corpus, permettant d'extraire les 1-grams (*hypersalé*), 2-grams (*hyper tonique*) et 3-grams (*hyper-inflation*) pour chacun des huit PHD étudiés. Les traitements effectués à ce stade sont les suivants : sélection des lignes commençant par l'un des préfixes et contenant l'une des dates sélectionnées pour l'étude (entre 1800 et 2010), élimination des chaînes contenant des caractères non alphabétiques, normalisation (espaces doubles, doubles tirets, mise en minuscule) et fusion des données pour les doublons.
- 20 **Étape 2.** Nous éliminons manuellement les formes fautives et non préfixées (étape 2a), ainsi que les formes préfixées non pertinentes (étape 2b), et récupérons les informations atomiques concernant le préfixe, la base de construction, le séparateur (fusion, trait d'union ou espace), ainsi que la fréquence par année. Les formes fautives sont essentiellement liées à des problèmes de numérisation des documents (ex. *superfonique* pour *supersonique*), et les formes non préfixées sont celles qui incluent les chaînes de caractères recherchées mais qui ne sont pas analysables morphologiquement (ex. *archipel*). Les formes préfixées non pertinentes sont celles dans lesquelles le préfixe n'indique pas le haut degré. Parmi celles-ci (cf. tableau 2), on trouve :
 - les mots qui indiquent une position spatiale supérieure (*architrave*, *hypermonde*, *supra-alvéolaire*),
 - les mots qui indiquent une position hiérarchique supérieure (*archilexème*, *super-intendant*, *supra-biologique*),
 - les mots qui indiquent une localisation externe (*ultramontain*, *extra-constitutionnel*),
 - les mots qui spécifient une unité de mesure (*mégabits*, *mégaparsec*),
 - les mots qui indiquent des regroupements supérieurs d'entités du type nommé par la base (*super-catégorie*, *supergalaxie*, *superstructure*).
- 21 Les formes présentant une ambiguïté entre une interprétation de haut degré et l'une des interprétations mentionnées ci-dessus sont exclues par défaut (*extra-nerveux*, *supra-sensible*).
- 22 **Étape 3.** Nous calculons la fréquence relative de chaque occurrence relevée, afin d'éviter les effets liés aux variations de la taille du corpus selon les années, en rapportant la fréquence absolue au nombre total d'occurrences par année, normalisée sur un million. À cette étape, nous ajoutons également la ou les parties du discours possibles pour chaque lexème. GNG comprend pour chaque entrée une information concernant la catégorie grammaticale, mais nos premiers échantillonnages ont montré que ces étiquettes, pour le français, étaient largement erronées. Pour surmonter cette difficulté, nous avons généré les informations de parties du discours en nous basant sur le dictionnaire morphosyntaxique du français moderne de Grezka *et al.* (2015).
- 23 **Étape 4.** Les données résultant des étapes antérieures comprennent un certain nombre de cas aberrants (*outliers*) concernant la fréquence d'emploi, et dont les propriétés biaisent les résultats globaux, notamment du point de vue des évolutions temporelles. Pour filtrer ces cas aberrants, nous avons éliminé toutes les formes ayant une fréquence inférieure ou supérieure à 2,5 fois l'écart-type⁸.

- 24 Les données quantitatives résultant de ces différents traitements sont détaillées dans le tableau 3. Au final, 3372 formes distinctes sont retenues (pour un total de 1 238 999 occurrences).

Tab. 3 : Données obtenues après prétraitement des résultats de *Google Ngram*

	Etape 1	Formes fautives et non préfixées	Etape 2a	Formes non pertinentes	Etape 2b/3	Etape 4
1-grams	7356	3923	3433	1410	2023	1947
2-grams	2579	1950	629	272	357	357
3-grams	2463	98	2365	1286	1079	1068
Total	12398	5971	6427	2968	3459	3372

2.2. Données résultantes

- 25 Il est possible d'effectuer une première analyse (non diachronique) des données extraites au moyen de GNG et prétraitées suivant la méthode décrite dans la sous-section précédente. Nous examinons pour chaque PHD le nombre de constructions distinctes recueillies, la forme graphique privilégiée, et la catégorie grammaticale des bases sélectionnées.
- 26 Le tableau 4 présente la distribution des formations lexicales et des occurrences selon les différents préfixes. Il apparaît que trois préfixes prévalent nettement dans l'expression du haut degré : *hyper-*, *ultra-* et *super-*, *hyper-* étant lui-même largement dominant dans les données recueillies. Ces trois préfixes sont suivis d'un deuxième groupe constitué de *extra-*, *archi-* et *méga-*. Enfin, *supra-* et *maxi-* se voient associer un nombre de formes et d'occurrences très faible dans le corpus exploité.

Tab. 4 : Formations lexicales et occurrences par préfixe après prétraitement des données

	Etape 1	Formes fautives et non préfixées	Etape 2a	Formes non pertinentes	Etape 2b/3	Etape 4
1-grams	7356	3923	3433	1410	2023	1947
2-grams	2579	1950	629	272	357	357
3-grams	2463	98	2365	1286	1079	1068
Total	12398	5971	6427	2968	3459	3372

- 27 Le tableau 5 présente les modes de liaison graphique associés à chaque PHD. La soudure domine largement, suivie du trait d'union et de l'espace, avec toutefois des différences importantes entre PHD. La distribution des formes graphiques est en effet significativement dépendante du type de préfixe (χ^2 (14, N = 3372) = 857,572, $p = 2e-16$). Certains préfixes comme *hyper-* ou *méga-* privilégient nettement la soudure graphique, tandis que d'autres, comme *ultra-* et *extra-*, favorisent le trait d'union, et que d'autres

encore, comme *archi-*, ont une distribution plus équilibrée entre les deux formes. Les formes espacées ne sont jamais dominantes, mais sont le plus représentées pour *super-*, *extra-* et *ultra-*. Ces variations du mode de liaison restent difficiles à interpréter : on ignore notamment si le trait d'union indique une simple décomposition morphémique ou une autonomisation plus avancée du préfixe. On peut néanmoins voir dans les différences de rapport observées un indice d'intégration affixale élevée pour des formes comme *hyper-* et *méga-*, et une pente vers la lexicalisation plus marquée pour des préfixes comme *super-* et *extra-*.

Tab. 5 : Modes de liaison par préfixe après prétraitement des données (% par préfixe)

	Archi	Extra	Hyper	Maxi	Méga	Super	Supra	Ultra	Total
Soudure	61	36	1137	4	92	373	10	234	1947
Trait d'union	62	79	172	6	22	241	8	478	1068
Espace	4	45	58	2	3	130	1	114	357
Total	127	160	1367	12	117	744	19	826	3372

- 28 Le tableau 6 présente les principales catégories grammaticales de base sélectionnées par les différents PHD dans les formations lexicales, d'après l'annotation automatique effectuée à l'étape 3 du prétraitement⁹. Globalement, les étiquettes N et Adj N représentent chacune environ 30% des résultats, suivies des étiquettes Adj (12%) et Adj_V (5%). On peut s'interroger sur l'existence de disparités dans les prédilections de sélection des différents PHD. Il semblerait par exemple qu'*hyper-* favorise des bases nominales, tandis qu'*archi-* sélectionne préférentiellement des bases adjectivales. Ce point nécessite toutefois un examen approfondi à partir des données de corpus.

Tab. 6 : Catégories grammaticales des bases morphologiques après prétraitement des données

Archi	Extra	Hyper	Maxi	Méga	Super	Supra	Ultra	Total
Adj_N : 51 V_Adj : 23 Adj : 15	Adj_N : 76 Adj_N_V : 37 Adj : 29	N : 503 Adj_N : 213 Adj : 194	N : 8 Adj_N : 2 Adv_N : 1	N : 56 V_N : 25 V_N_Adj : 3	N : 320 Adj_N : 184 V_N : 83	Adj_N : 8 Adj : 6 N : 4	Adj_N : 435 Adj : 131 N : 90	N : 997 Adj_N : 972 Adj : 417 V_Adj : 193

- 29 Ajoutons pour terminer que la distribution des fréquences d'emploi par préfixe répond à la courbe de Zipf caractéristique à la fois du vocabulaire dans son ensemble et des formants productifs du lexique. Le modèle de distribution des formes construites renvoie à une organisation du lexique en deux groupes principaux : une partie stabilisée, comprenant des items de fréquence moyenne, forte ou très forte, et constituant entre 80% et 90% du vocabulaire actif, et une partie plus instable, comprenant des items de fréquence faible ou très faible, et constituant entre 5% et 20% du vocabulaire actif (cf. Zipf, 1965 [1935]; Mandelbrot, 1953; Baayen, 2001). En moyenne, pour les huit PHD que nous étudions, cette frange instable représente entre

8% et 11% du vocabulaire actif, ce qui semble indiquer un niveau de productivité comparable pour l'ensemble de ces PHD.

2.3. Propriétés étudiées et analyse diachronique

- 30 Pour étudier les caractéristiques et l'évolution des différentes formations lexicales, l'idéal consisterait à prendre en compte un maximum de paramètres linguistiques (forme de la construction, catégorie grammaticale de la base, classe sémantique, constructions syntaxiques) et extralinguistiques (propriétés des contextes). Dans le cadre de cette étude, étant donné les informations disponibles dans le corpus utilisé, nous exploiterons les propriétés suivantes : fréquence absolue et relative, forme de la base, mode de liaison préfixe-base, préfixe, forme construite complète.
- 31 L'observation de la fréquence d'emploi des formes comportant un PHD permet d'évaluer l'implantation cognitive (*entrenchment*, cf. Langacker, 1987 ; Blumenthal-Dramé, 2012 ; Schmid, 2007, 2020 ; *inter alia*) et la diffusion des lexèmes dans la communauté linguistique. GNG nous fournit des informations de fréquence absolue par année, ainsi que le nombre total de mots par année. La fréquence relative peut donc être calculée d'après le rapport entre le nombre d'occurrences d'une forme cible et le nombre total d'occurrences dans le corpus découpé par tranches annuelles, ce qui permet de mitiger les variations de volume du corpus par année¹⁰. La fréquence relative peut ainsi directement être utilisée pour comparer les évolutions dans le temps des différentes propriétés étudiées. Ces fréquences permettent également, lorsqu'elles sont appliquées à une ou plusieurs propriétés des formes construites, d'obtenir une estimation des distributions entre les valeurs et leur évolution. L'étude quantitative nous permettra ainsi non seulement d'évaluer l'importance relative de chacun des préfixes dans l'expression du haut degré, mais également d'identifier des phases saillantes dans l'évolution de leurs emplois et de leurs modes de construction.
- 32 La productivité des PHD est une autre caractéristique qui sera étudiée ici. Les PHD partagent en effet avec d'autres morphèmes la capacité de construire de nouveaux lexèmes, par concaténation avec de nouvelles bases. Pour évaluer quantitativement cette propriété, Baayen (1992, 2009) a proposé de distinguer entre productivité réalisée, productivité en expansion et productivité potentielle. Cependant, selon Cartier (2018), seule la productivité en expansion¹¹ est effectivement mesurable. On peut néanmoins calculer, de façon globale, le rapport entre les lexèmes construits de faible fréquence et la totalité des lexèmes construits attestés. Il apparaît que pour les suffixes du français contemporain, la distribution des fréquences respecte une courbe log-normale, avec un plus ou moins grand nombre de formations de faible fréquence et, selon l'historique du formant, une plus ou moins longue traîne de formations de haute ou très haute fréquence (cf. Cartier, 2018). Cette approche synchronique sera complétée ici par l'étude, pour chaque PHD, de la distribution entre les formes qui apparaissent, celles qui disparaissent et celles qui sont préservées au cours d'une période donnée, et par l'étude de l'évolution de cette distribution dans le temps¹².

3. Résultats et analyses¹³

- 33 Dans cette section, nous présentons les données recueillies concernant la fréquence d'emploi des PHD dans le GNG, pour la période 1800-2010. Nous exposons d'abord les

principales tendances observées pour l'ensemble des PHD, puis nous analysons le détail des évolutions par préfixe. Enfin, nous examinons les cas de concurrence avérée entre les mots préfixés par un PHD dérivés des mêmes bases lexicales.

3.1. Tendances générales et comparées

- 34 Au cours des 19^e et 20^e siècles, l'emploi des PHD connaît une croissance continue dans le corpus étudié, comme le montre la figure 1 ci-dessous¹⁴. On peut distinguer différentes phases dans cette progression globale : une croissance faible et régulière jusqu'au début des années 1920 (avec une légère accentuation entre 1890 et 1920), puis plus soutenue dans l'entre-deux-guerres, avant une stagnation générale entre 1944 et 1995 (et même une phase de recul entre 1975 et 1984), et enfin une augmentation de la fréquence d'emploi jusqu'en 2010. Ces tendances générales peuvent évidemment masquer des disparités importantes entre les différents PHD, ou être influencées par le comportement spécifique de certains PHD fortement représentés dans les données.

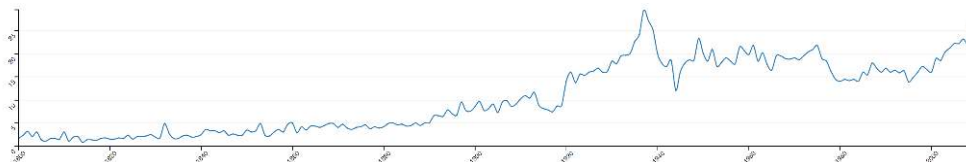


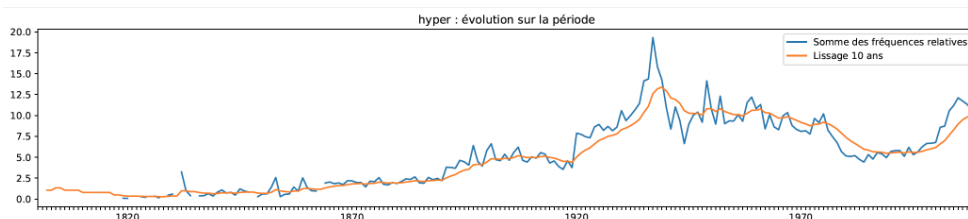
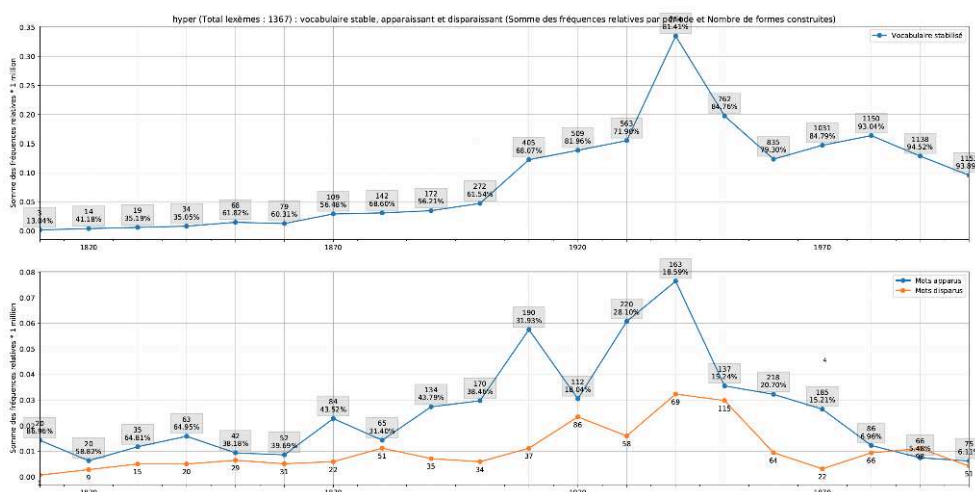
Fig. 1 : Fréquence d'emploi relative de l'ensemble des formes comprenant un PHD (1800-2010)

- 35 L'évolution comparée des différents PHD présente en effet une situation complexe, dont nous pouvons indiquer schématiquement les phases principales¹⁵. Comme nous l'avons vu dans la section 2.2., *hyper-*, *ultra-* et *super-* sont les PHD les plus fréquemment employés entre 1800 et 2010, mais leurs rapports évoluent au fil du temps. Au début de la période, *super-* domine nettement. En seconde position, *hyper-* se détache des autres PHD, qui sont eux très faiblement représentés. La fréquence d'emploi de *hyper-*, mais aussi d'*ultra-* croît alors progressivement jusqu'à devenir équivalente à celle de *super-* vers 1835. Cette situation perdure jusqu'à 1865. L'emploi d'*hyper-* augmente alors significativement et domine nettement jusqu'à aujourd'hui (avec toutefois un rapprochement de *super-* et *ultra-* entre 1980 et 2000). De son côté, l'emploi d'*ultra-* et de *super-* stagne jusqu'aux environs de 1905 pour *ultra-* et 1920 pour *super-*, avant de connaître une nette augmentation. Les deux préfixes se rejoignent vers 1940, puis s'emploient avec la même fréquence jusqu'en 1975. Il leur correspond ensuite des périodes d'augmentation et de diminution complémentaires (*super-* domine de 1975 à 1990, puis vient le tour d'*ultra-* jusqu'en 2010).
- 36 L'emploi des 5 autres PHD est très mineur, sinon inexistant, dans la période 1800-1870. À partir de 1870, l'emploi d'*extra-*, qui avait déjà donné quelques signes d'accroissement, se répand et domine celui des 4 autres PHD, jusqu'à un pic de fréquence vers 1940. Au cours de la même période, la fréquence d'emploi de *méga-* croît également, mais plus lentement que celle d'*extra-*, avant que les deux courbes ne se rejoignent autour de 1940. La tendance générale d'*extra-* est ensuite de décroître, et *extra-* et *méga-* sont dans des zones de fréquence proches entre 1940 et 1980. *Méga-* domine légèrement dans cette période puis plus nettement jusqu'en 2010. De son côté, l'emploi d'*archi-* augmente très progressivement sur l'ensemble de la période et côtoie des zones de fréquence similaires à celles d'*extra-* à partir de 1970 environ (avec une légère domination depuis 1990). Quant à *maxi-* et *supra-*, leur emploi en tant que PHD

reste dans l'ensemble très faible, avec néanmoins quelques poussées de *supra-* à partir de 1920, qui vient sporadiquement concurrencer *archi-* et *extra-*.

3.2. Détail des évolutions

- 37 Nous examinons à présent le détail des évolutions spécifiques, en nous concentrant sur la fréquence d'emploi et la productivité de chaque PHD considéré individuellement. Le PHD le plus fréquemment employé dans la période étudiée, à savoir *hyper-*, a une évolution globale proche de celle décrite ci-dessus pour les PHD considérés conjointement, sa haute fréquence relativement aux autres PHD déterminant en grande partie les observations de tendances générales (cf. figure 2). L'emploi d'*hyper-*, faible en début de période, croît d'abord doucement jusqu'en 1890, puis de plus en plus rapidement durant deux phases (1890-1920 et 1920-1940), avant de stagner (1940-1975), de décroître et de se maintenir à un niveau plus bas (1975-2000), pour enfin augmenter à nouveau (2000-2010). Comme l'indique le tableau 5, *hyper-* privilégie très nettement la soudure graphique, et on n'observe pas d'évolution notable dans le relevé diachronique des fréquences d'emploi selon la forme graphique. Les bases de construction d'*hyper-* restent très majoritairement des noms et des adjectifs sur l'ensemble de la période. Un examen qualitatif des formes construites fait apparaître l'abondance d'emploi dans des domaines de spécialité (biologie, médecine), clairement corrélé aux formes graphiques soudées, dans lesquelles l'emploi prototypique d'*hyper-* est préfixal et formellement intégré aux mots construits. Pour compléter l'analyse, on peut examiner l'évolution dans le temps de la répartition entre les formes qui apparaissent, celles qui disparaissent et celles qui sont conservées au fil du temps. C'est l'objet de la figure 3, dans laquelle la courbe supérieure correspond au vocabulaire stabilisé, et les deux courbes inférieures aux formes qui apparaissent et qui disparaissent au cours de chaque période successive de dix ans. Dans le cas des formes nouvelles et stabilisées, nous indiquons pour chaque période le nombre d'items concernés et le pourcentage du vocabulaire actif correspondant (i.e. la somme des formes stabilisées et des formes apparues dans la période), ce qui permet d'approcher la productivité affixale en diachronie. Les courbes sont par ailleurs établies en rapport avec le nombre total de mots présents dans le corpus à chaque période, pour pondérer le relevé en fonction de l'évolution du volume du corpus. Dans le cas d'*hyper-*, on constate d'emblée que l'évolution de la fréquence d'emploi relative du vocabulaire stabilisé est en phase avec l'évolution globale du PHD, excepté en fin de période : on observe une progression constante et régulière jusque dans les années 1940, puis une phase de décroissance. L'apparition de formes nouvelles constitue l'élément saillant de cette évolution, avec des taux de renouvellement compris entre 18% et 86% jusqu'en 1940, et une décline continue au fur et à mesure de l'augmentation du vocabulaire stabilisé. Le renouvellement lexical est également marqué par le nombre important de formes disparues entre 1920 et 1950. On observe une baisse progressive du taux de mots nouveaux à partir de 1950, ce qui n'entrave toutefois pas l'augmentation de la fréquence d'emploi du préfixe en fin de période (certains items stabilisés voyant alors leur fréquence d'emploi augmenter).

Fig. 2 : Fréquence d'emploi relative des formes *hyper+X* (1800-2010)¹⁶Fig. 3 : Courbes d'évolution des formes *hyper+X* stables, apparues et disparues (1800-2010)

- 38 L'emploi d'*ultra-*, quasiment inexistant au début de la période, suit à partir de 1815 et jusqu'en 1870 environ une croissance assez régulière, avant une phase de stagnation de 1865 à 1905, et une phase de développement plus accru jusqu'en 2010 (avec néanmoins un affaissement temporaire entre 1960 et 1985), comme le montre la figure 4. Graphiquement, le trait d'union domine, tant dans les types construits que dans les occurrences observées en corpus, mais il est plus régulièrement concurrencé par les formes fusionnées à partir de 1905. Les bases ambiguës entre adjectif et nom prévalent dans la construction avec *ultra-* sur l'ensemble de la période, en fréquence d'emploi des items, avec une nette croissance et une domination accentuée après 1945. Selon Rey (2016), l'évolution lexicale des constructions en *ultra-* est profondément modifiée au moment de la Révolution française: alors qu'*ultra-* s'employait essentiellement précédemment dans des expressions issues du latin, on voit apparaître en 1792 les mots *ultra-révolution(naire)* puis *ultra-royaliste*, qui dès le début du 19^e siècle donneront lieu par analogie à une abondance de néologismes dans le domaine politique et religieux (ex. *ultra-patriotique*, *ultra-républicain*, *ultra-catholique*), le préfixe s'appliquant à partir des années 1830 à toute propriété poussée à l'extrême (ex. *ultra-romantique*, *ultra-comique*, *ultra-moderne*). D'une manière générale, le renouvellement des formes dans le cas d'*ultra-* évolue de manière proche de celle observée pour *hyper-*, comme on le voit dans la figure 5, avec jusqu'aux années 1940 une croissance régulière du nombre de constructions nouvelles et stabilisées, qui suit la courbe d'évolution de la fréquence d'emploi du préfixe, puis une relative stagnation, voire une baisse, qui n'entrave pas l'augmentation générale de la fréquence d'emploi du préfixe. Le taux de formes nouvellement créées, moins important en moyenne que dans le cas d'*hyper-*, évolue selon trois phases : une forte contribution à la croissance du préfixe entre 1800 et 1840 (> 55%), puis un taux oscillant entre 20% et 40% de 1850 à 1970, et enfin une chute à

moins de 10% entre 1980 et 2010. La situation en fin de période est celle d'un lexique stabilisé à environ 95%, mais dont la fréquence d'emploi continue d'augmenter.

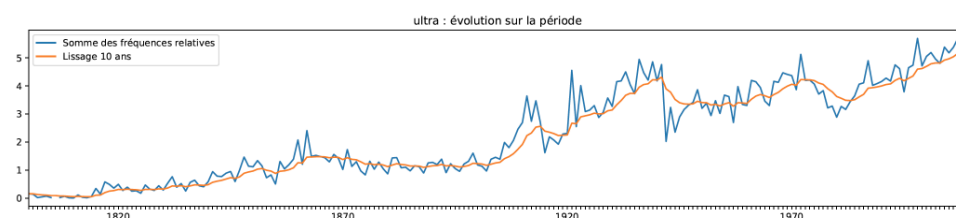


Fig. 4 : Fréquence d'emploi relative des formes *ultra+X* (1800-2010)

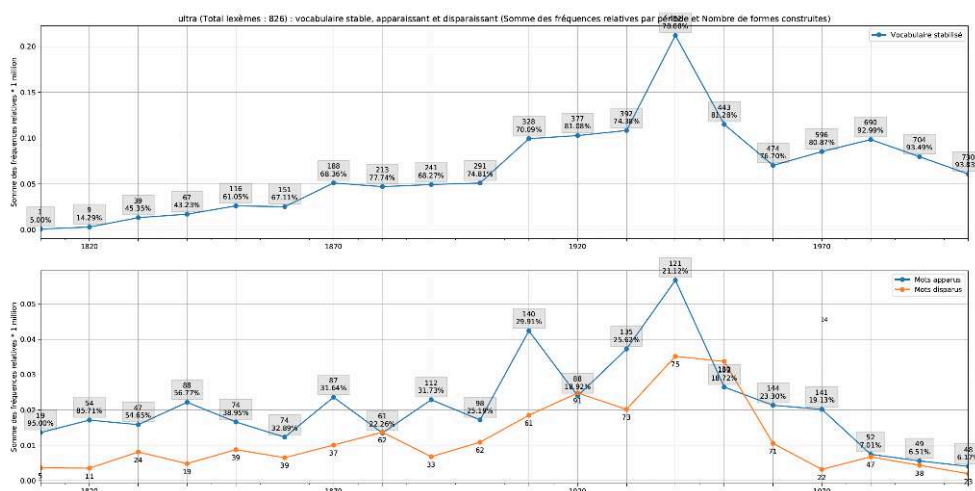
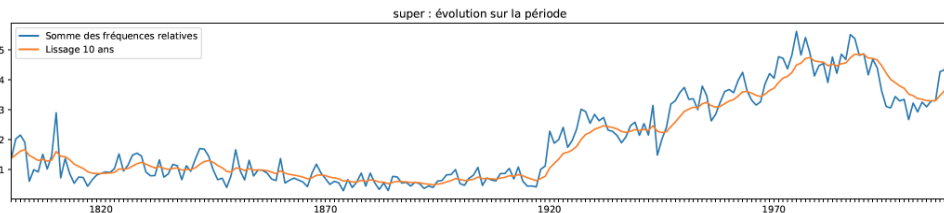
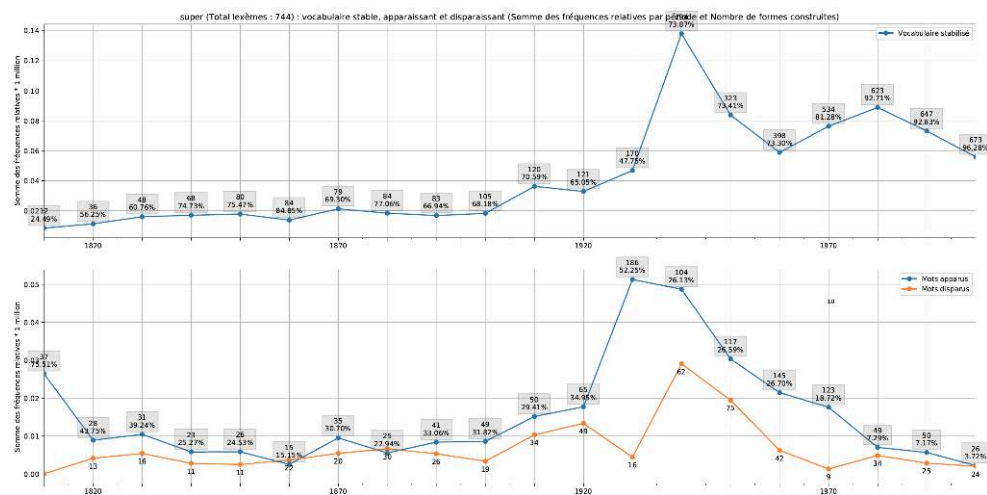
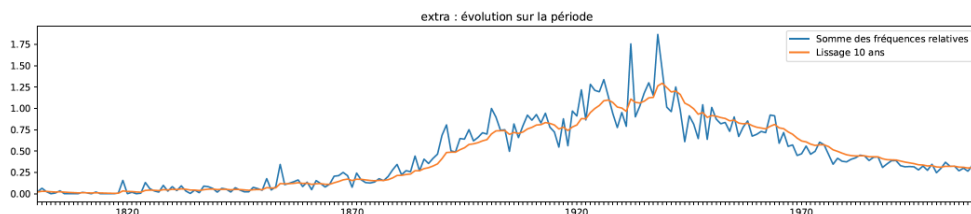


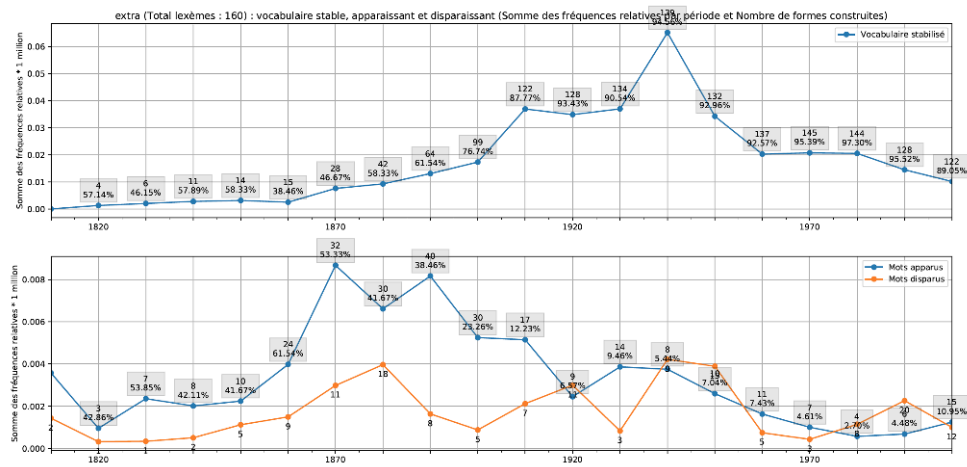
Fig. 5 : Courbes d'évolution des formes *ultra+X* stables, apparues et disparues (1800-2010)

- 39 L'emploi de *super-*, plus élevé que celui des autres PHD en début de période, stagne et même baisse légèrement entre 1800 et 1920, puis croît de manière très importante jusqu'aux environs de 1990, avant de connaître une nouvelle baisse et une remontée en toute fin de période (cf. figure 6). Graphiquement, les formes soudées dominent, même si la fréquence d'emploi des occurrences comportant un trait d'union s'accroît entre 1940 et 1980. Les emplois avec espace augmentent très progressivement à partir de 1950, et connaissent une hausse importante en fin de période, ce qui constitue un indice possible de lexicalisation (comme adverbe), et est en partie responsable du regain d'emploi tardif. *Super-* se construit principalement à partir de bases nominales et, dans une moindre mesure, de bases ambiguës entre nom et adjectif, cette prédilection s'accroissant à mesure qu'augmente la fréquence d'emploi du préfixe. Le nombre de constructions existantes est à peu près stable jusqu'en 1920. Il augmente substantiellement entre 1920 et 1940, avant de subir une baisse générale. Le taux de mots nouveaux oscille entre 15% et 50% sur l'ensemble de la période 1800-1960, avec des valeurs généralement aux alentours de 20% ou 30%. Il baisse ensuite de manière importante entre 1970 et 2010 (jusqu'à 3%). Le vocabulaire semble donc se stabiliser à partir des années 1960, ce qui n'empêche pas que s'accroisse la fréquence d'emploi générale des formes en *super-*.

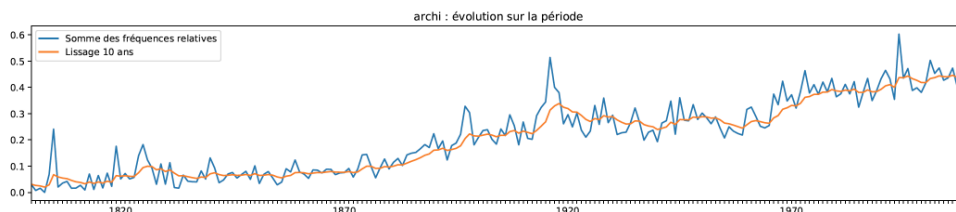
Fig. 6 : Fréquence d'emploi relative des formes *super+X* (1800-2010)Fig. 7 : Courbes d'évolution des formes *super+X* stables, apparues et disparues (1800-2010)

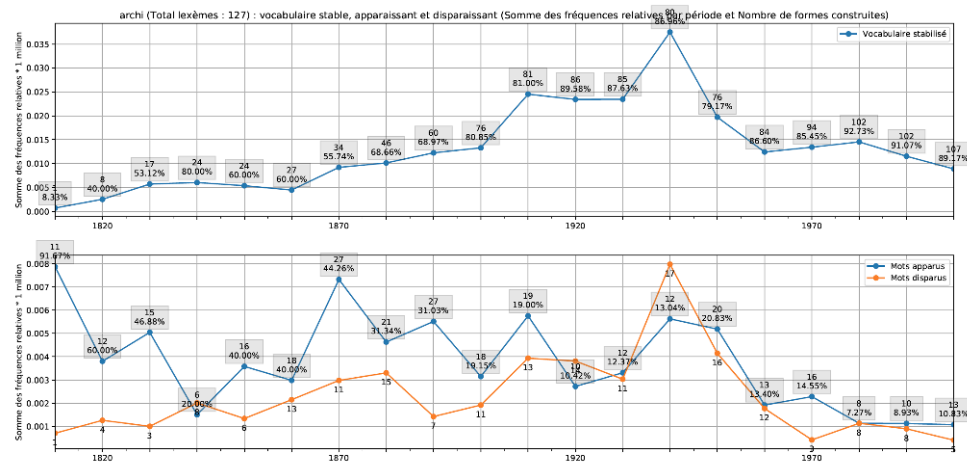
- 40 **Extra-** dans ses emplois comme PHD présente la particularité, après une phase initiale de basse fréquence d'emploi, d'augmenter régulièrement entre 1850 et 1940, puis de décroître tout aussi régulièrement jusqu'en 2010. Graphiquement, le trait d'union domine sur l'ensemble de la période, à la fois dans les formes construites et dans la fréquence d'emploi en corpus. Les bases de prédilection restent très majoritairement les formes ambiguës entre nom et adjectif. Contrairement à ce qu'on observe dans le cas de *hyper-* et *ultra-*, les bases de construction de *extra-* appartiennent dès le départ à la langue générale (ex. *extra-fin*, *extra-frais*, *extra-pur*, *extra-blanc*). La figure 9 précise certaines évolutions : tandis que le nombre de formes attestées se calque sur la progression fréquentielle, le taux de formes nouvelles évolue de manière contrastée. Il se situe dans une fourchette de 40% à 60% du vocabulaire actif entre 1800 et 1890, mais chute ensuite pour s'établir à des niveaux très faibles (moins de 10% à partir de 1920), accompagnant la décroissance globale des emplois du préfixe.

Fig. 8 : Fréquence d'emploi relative des formes *extra+X* (1800-2010)

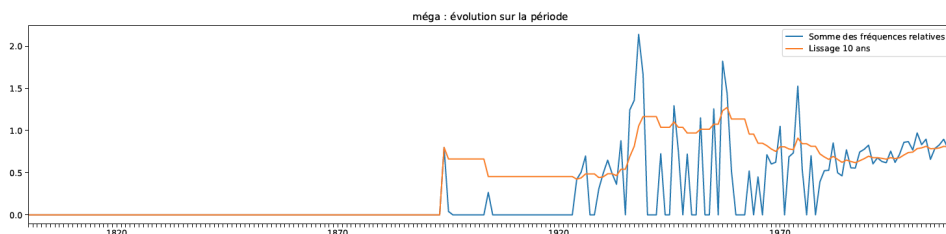
Fig. 9 : Courbes d'évolution des formes *extra+X* stables, apparues et disparues (1800-2010)

- 41 L'emploi d'*archi-* pour exprimer le haut degré est stable entre 1800 et 1880, puis il connaît une augmentation constante jusqu'en 2010. Tout au long de cette évolution, la fréquence d'emploi des formes fusionnées est équivalente à celle des formes avec trait d'union, l'espace étant très peu utilisé. Bien que limitées en nombre, certaines bases se construisent avec *archi-* dès le début de période pour former des dérivés de haut degré (ex. *archi-comble*, *archi-grand*, *archi-faux*). Les bases les plus sollicitées sont des adjectifs dans toutes les phases de l'évolution, la construction avec les noms étant plus rare que pour des PHD comme *hyper-* ou *super-*. Certains facteurs de l'évolution diachronique d'*archi-* apparaissent dans la figure 11. On constate une augmentation régulière de l'ensemble du lexique actif jusque dans les années 1940, puis une décline accompagnée d'une augmentation du taux de vocabulaire stabilisé. De fait, le taux de renouvellement lexical est beaucoup plus important en début de période, notamment entre 1800 et 1890, qu'en fin de période ^{F0}₂₀ s'établissant sous la barre des 10% à partir de 1980.

Fig. 10 : Fréquence d'emploi relative des formes *archi+X* (1800-2010)

Fig. 11 : Courbes d'évolution des formes *archi+X* stables, apparues et disparues (1800-2010)

- 42 **Méga-** émerge véritablement comme PHD après 1870, date à partir de laquelle son emploi augmente régulièrement, avec quelques pics entre 1940 et 1980, comme on le voit dans la figure 12. Les formes les plus fréquemment employées sont des formes fusionnées et sont construites sur base nominale. On peut noter que dans la plupart de ses emplois, *méga-* porte sur la taille plutôt qu'à proprement parler sur la haute intensité (ex. *méga-concert*, *mégastructure*). Par ailleurs, le renouvellement lexical paraît plus fluctuant que dans le cas des autres PHD. Si le taux de vocabulaire stabilisé tend généralement à l'augmentation, il ne se fixe pas à un très haut niveau en fin de période, contrairement à ce qu'on pouvait observer pour *hyper-*, *ultra-*, *super-*, *extra-* et *archi-*. Le taux de formes nouvelles en *méga-* reste très variable entre 1960 et 2010, ce qui laisse penser que la structuration du lexique correspondant n'est pas stabilisée. Notons toutefois que le nombre peu important de bases prises en compte ici (117) pourrait biaiser les résultats observés.

Fig. 12 : Fréquence d'emploi relative des formes *méga+X* (1800-2010)

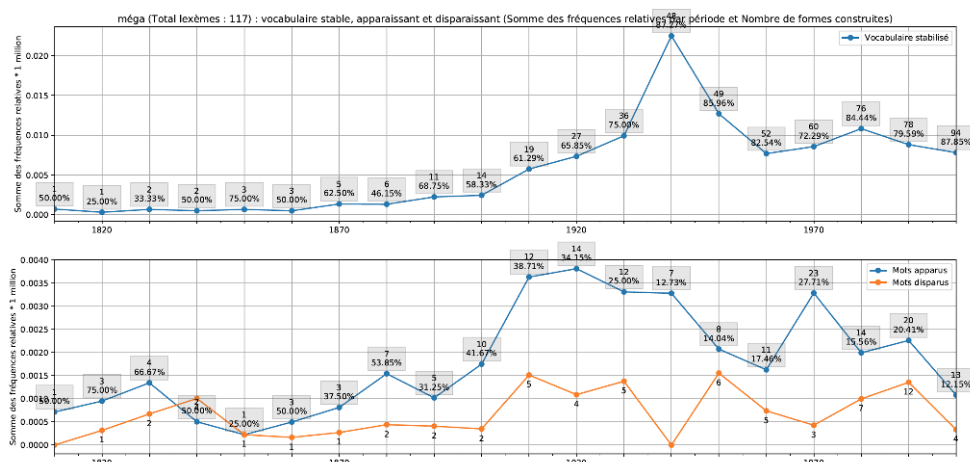


Fig. 13 : Courbes d'évolution des formes *méga+X* stables, apparues et disparues (1800-2010)

- 43 **Supra-** et **maxi-**, enfin, se concentrent dans notre corpus sur quelques formes seulement. *Supra-* comme PHD n'apparaît véritablement qu'après 1920 et sa fréquence d'emploi se maintient à un niveau bas. Les formes les plus fréquemment employées sont fusionnées et sont construites à partir de bases ambiguës entre adjectif et nom. L'emploi de *maxi-* comme PHD est quasi-inexistant dans le corpus analysé (12 formes pour 2122 occurrences). Dans les deux cas, le faible nombre de bases de construction (13 pour *supra-* et 10 pour *maxi-*) rendent difficile l'évaluation des tendances concernant la productivité du préfixe.

3.3. Concurrence lexicale

- 44 En complément de l'évolution comparée de l'emploi et de la productivité des différents PHD, il est intéressant d'examiner la concurrence lexicale directe entre PHD, i.e. la coexistence possible de formes construites à partir d'une même base lexicale avec des PHD variés. Des situations de concurrence sont en effet régulièrement observées entre deux PHD ou plus, comme dans le cas des formes listées sous (7).
- (7) a. archi+moderne, extra+moderne, hyper+moderne, super+moderne, ultra+moderne
 b. extra+dense, hyper+dense, super+dense, ultra+dense
 c. hyper+patriotisme, super+patriotisme, ultra+patriotisme
 d. hyper+stable, ultra+stable
- 45 On recense plus exactement 1 cas de concurrence entre 5 PHD, 11 cas de concurrence entre 4 PHD, 42 cas de concurrence entre 3 PHD, et 225 cas de concurrence entre 2 PHD. Les concurrences deux à deux sont comptabilisées dans le tableau 7, sans tenir compte des variations de forme graphique, mais en intégrant les expressions très fréquentes (*outliers*) exclues dans les analyses précédentes (cf. section 2.1.).

Tab. 7 : Concurrence des PHD dans la réalisation lexicale (bases partagées pour la période 1800-2010, indépendamment de la variation graphique). Les pourcentages de types concernés par PHD sont indiqués en ligne¹⁷.

	Archi	Extra	Hyper	Maxi	Méga	Super	Supra	Ultra
Archi	-	3 (3,5%)	3 (3,5%)	0	0	6 (7,0%)	0	13 (15,1%)
Extra	3 (3,3%)	-	25 (27,7%)	0	0	26 (28,8%)	1 (1,1%)	44 (48,9%)
Hyper	3 (0,2%)	25 (2,1%)	-	1 (0,08%)	7 (0,6%)	104 (8,6%)	0	111 (9,2%)
Maxi	0	0	1 (10%)	-	0	2 (20%)	0	1 (10%)
Méga	0	0	7 (6,7%)	0	-	9 (8,7%)	0	1 (1,0%)
Super	6 (1,1%)	26 (4,9%)	104 (19,6%)	2 (0,4%)	9 (1,7%)	-	4 (0,8%)	65 (12,3%)
Supra	0	1 (7,7%)	0	0	0	4 (30,8%)	-	1 (7,7%)
Ultra	13 (2,5%)	44 (8,4%)	111 (21,2%)	1 (0,2%)	1 (0,2%)	65 (12,4%)	1 (0,2%)	-

- 46 On observe que les proportions de formes concurrentes par affixe sont globalement élevées, alors qu'on peut douter de l'existence de différences sémantiques entre les doublons en question. Autrement dit, la possibilité qu'il existe lexicalement des formes comprenant un PHD en surabondance paraît forte. Le taux de doublonnage est particulièrement important dans le cas d'*extra-*, ce qui peut être mis en relation avec le fait que ce préfixe s'emploie massivement pour exprimer l'extériorité, et que son existence dans la langue ne se justifie pas par l'exclusivité de l'expression du haut degré et de la formation de dérivés correspondants. À l'inverse, les préfixes comme *hyper-*, plus spécifiquement dédiés à la dérivation d'expressions de haut degré, semblent former plus de dérivés exclusifs, ce que l'on pourrait lier *in fine* à leur persistance en langue. Il reste que le constat général d'un taux de concurrence lexicale élevé entre les différents PHD doit être nuancé, d'une part, par le fait que les formes retenues dans GNG sont celles qui ont une fréquence supérieure ou égale à 40, ce qui exclut les formes peu fréquentes, dont on peut penser qu'elles sont moins exposées à la concurrence, et d'autre part, par le fait que la concurrence est évaluée sur l'intégralité de la période 1800-2010, et que potentiellement les doublons morphologiques ne coexistent pas ou peu au cours de périodes plus réduites, si leurs emplois se succèdent en diachronie.
- 47 Pour affiner l'analyse du doublonnage lexical entre PHD, il faudrait étudier précisément l'évolution des différents cas de concurrence existants. À titre exploratoire, nous avons examiné le cas le plus représenté, à savoir celui de la concurrence entre *hyper-* et *ultra-*, ces deux préfixes étant des PHD très fréquents, dont les différences sémantiques (ex. en termes d'intensité décrite) ne sont pas manifestes. Dans l'évolution des fréquences d'emploi des formes construites, parmi les 71 cas de concurrence stricte entre *hyper-* et *ultra-*, on peut distinguer schématiquement quatre cas de figure.
- 48 On remarque tout d'abord qu'il existe des cas de concurrence très fortement déséquilibrés, dans lesquels une forme qui peut connaître d'importantes variations de fréquence domine très nettement l'autre, qui au contraire reste cantonnée à un nombre

d'emplois très faible. L'évolution comparée d'*ultra+basique* et d'*hyper+basique*, illustrée dans la figure 14, constitue un exemple de ce type. Nous avons recensé 10 configurations de ce genre sur les 71 cas examinés. La forme dominante est *ultra-* dans 9 de ces cas et *hyper-* dans le dernier.

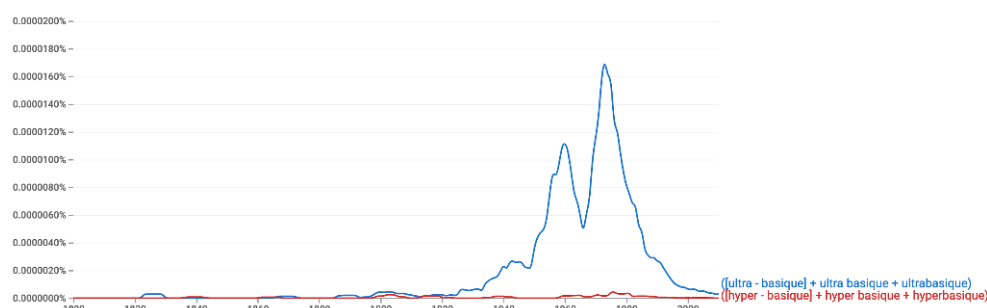


Fig. 14 : Évolution temporelle de l'emploi des formes *ultra+basique* et *hyper+basique* (lissage sur 3 ans)

- 49 Dans un second cas de figure, les formes concurrentes ont des évolutions parallèles, au sens où, même si une forme domine l'autre en fréquence, leur progression d'emploi dans le temps suit à peu près la même courbe temporelle. La concurrence entre *hyper+sophistiqué* et *ultra+sophistiqué* fournit un exemple de ce type (cf. figure 15). Nous avons recensé 16 cas d'évolutions parallèles parmi les 71 cas de concurrence observés.

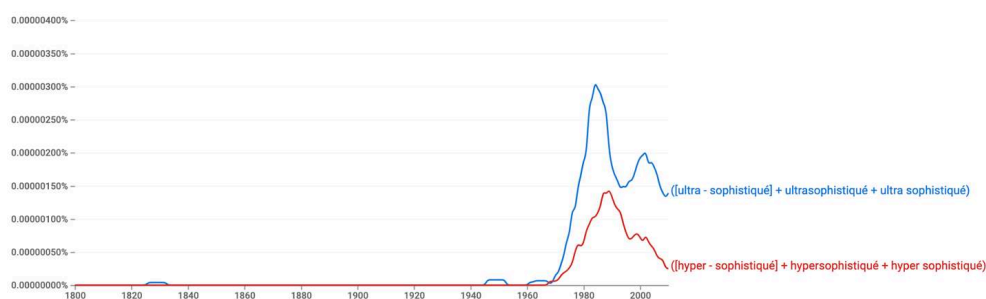


Fig. 15 : Évolution temporelle de l'emploi des formes *ultra+sophistiqué* et *hyper+sophistiqué* (lissage sur 3 ans)

- 50 Une troisième configuration possible est celle des évolutions complémentaires, dans lesquelles les courbes d'évolution des formes concurrentes sont inversées (avec éventuellement des phases d'alternance de forme dominante). La complémentarité absolue n'est jamais observée dans les cas que nous avons examinés, mais deux d'entre eux s'en rapprochent fortement, en l'occurrence *hyper+logique* / *ultra+logique* et *hyper+scientifique* / *ultra+scientifique*, ce dernier étant donné en exemple dans la figure 16.

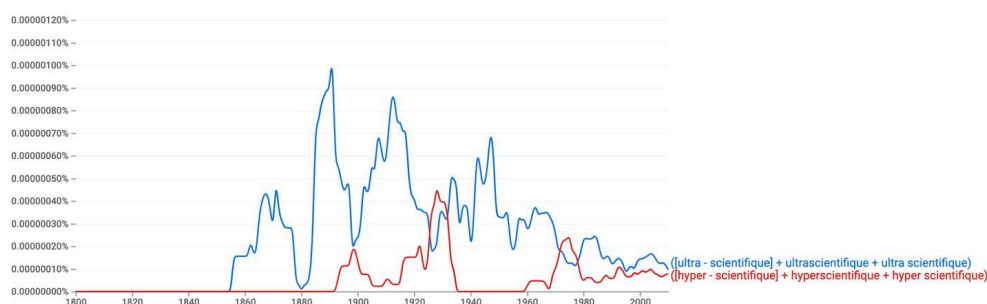


Fig. 16 : Évolution temporelle de l'emploi des formes *ultra+scientifique* et *hyper+scientifique* (lissage sur 3 ans)

- 51 Le quatrième et dernier cas de figure consiste en un schéma hybride, dans lequel les doublons alternent des phases d'évolution en parallèle et des phases d'évolution complémentaire (souvent avec inversion des formes dominantes). La comparaison d'*hyper+personnel* et d'*ultra+personnel* fournit un exemple de ce type. Comme on le voit dans la figure 17, les courbes d'emploi des deux expressions comportent des phases d'évolution inversée, par exemple en 1900-1925 ou en 1970-1985, mais aussi des phases d'évolution parallèle, par exemple en 1950-1970 ou en 1985-2010. D'après notre décompte, 43 cas de concurrence sur 71 présentent un profil hybride. Dans la plupart de ces cas, les phases de parallélisme sont dominantes.

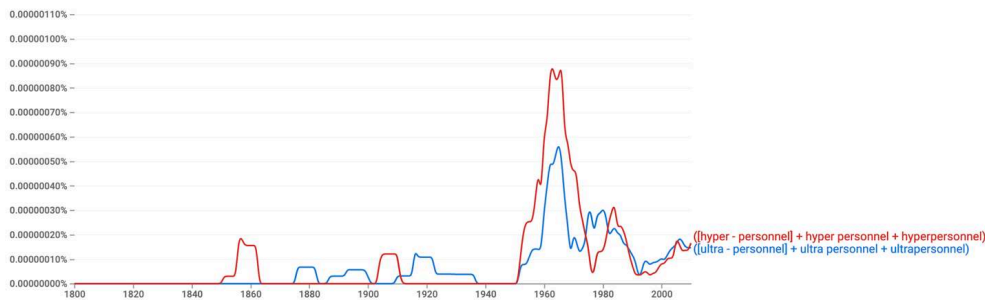


Fig. 17 : Évolution temporelle de l'emploi des formes *ultra+personnel* et *hyper+personnel* (lissage sur 3 ans)

- 52 Il ressort de ces observations que la plupart des doublons lexicaux préfixés par *hyper-* et *ultra-* n'évoluent pas dans le sens d'une réduction des formes en diachronie. Un dérivé domine généralement l'autre en fréquence d'emploi, mais il est assez rare qu'une forme supplante durablement une autre, alors qu'à l'inverse les doublons présentent souvent au cours du temps des phases d'évolution parallèle. L'examen des cas de concurrence entre *hyper-* et *ultra-* fait donc apparaître la vitalité de ces deux préfixes, notamment au cours du 20^e siècle. L'évolution en diachronie des doublons préfixés par des PHD devrait toutefois être examinée à plus grande échelle pour confirmer les premiers résultats présentés ici. Il serait également intéressant, lorsque le corpus intègre cette information, d'étudier la répartition des PHD par sous-genres, avec l'idée que certains types de textes favorisent l'emploi de PHD en particulier.

Conclusion

- 53 Dans cet article, nous avons étudié l'évolution de l'emploi de huit préfixes de haut degré du français entre 1800 et 2010. D'une manière générale, les PHD voient leur emploi augmenter de manière significative au cours de la période, notamment entre 1900 et 1970. Un examen détaillé des fréquences d'emploi par préfixe fait toutefois apparaître des situations nuancées, avec des phases d'augmentation plus (ex. *archi-*) ou moins (ex. *super-*) précoces, un éventuel tassement en fin de période (ex. *extra-*), et une productivité plus ou moins importante selon les cas. Trois PHD (*hyper-*, *ultra-*, *super-*) dominant nettement, aussi bien dans le nombre de formes construites que dans la fréquence d'emploi, tandis que d'autres (*maxi-*, *supra-*) sont très peu représentés dans le corpus. Malgré ces disparités, la situation de surabondance affixale, dans le cadre de l'expression du haut degré, est avérée et persistante en diachronie, comme le montre le développement parallèle de doublons morphologiques peu discernables sémantiquement. On n'assiste notamment pas au recul d'*hyper-*, *ultra-* ou *super-* face aux

deux autres ⁵⁰ les trois PHD connaissant même une recrudescence d'emplois en fin de période ⁵¹ et des cas de synonymie durables sont observés pour des formes concurrentes construites à l'aide de ces trois PHD. La concurrence entre PHD ne se résout donc pas en diachronie, alors qu'on peut s'interroger sur l'apparition de nouveaux préfixes destinés à exprimer le haut degré (ex. *hypra-*, qui apparaît comme un amalgame de *hyper-* et *supra-*).

- 54 Il est important de rappeler que les résultats présentés ici sont tributaires du corpus analysé qui, bien qu'il intègre des productions variées, reste fondamentalement un corpus écrit et livresque. L'étude des PHD devrait être complétée par l'examen de leur emploi dans des genres plus diversifiés, en particulier à l'oral ou dans des registres plus informels. De même, une étude en micro-diachronie, sur des données recueillies après 2010, pourrait utilement compléter la description du phénomène. Le recours à des corpus ou à des bases de données mieux contrôlés que *Google Books* (ex. *Timestamped JSI web corpus*, *Néoveille*), permettant d'analyser un nombre plus important de propriétés linguistiques (notamment les contextes d'emploi et les structures distributionnelles) et extra-linguistiques (types de texte, origine géographique, données socio-pragmatiques), avec une prise en compte des nouvelles formations dès la première occurrence, devrait conduire à des résultats plus fins concernant l'emploi des PHD en français contemporain, leur productivité respective, et le dynamisme de leur concurrence.

BIBLIOGRAPHIE

- AMIOT, D., 2004, « Haut degré et préfixation », *Travaux Linguistiques du Cerlico*, 17, p. 91-104.
- ARNDT-LAPPE, S., 2014, "Analogy in suffix rivalry: the case of English *-ity* and *-ness*", *English Language and Linguistics*, 18, p. 497-548.
- ARONOFF, M., 2016, "Competition and the lexicon", dans A. Elia, C. Iacobini, M. Voghera (éds), *Livelli di Analisi e Fenomeni di Interfaccia*, Roma, Bulzoni, p. 39-52.
- BONAMI, O. & STRNADOVÁ, J., 2019, "Paradigm structure and predictability in derivational morphology", *Morphology*, 29(2), p. 167-197.
- BONAMI, O. & THUILLIER, J., 2019, "A statistical approach to rivalry in lexeme formation: French *-iser* and *-ifier*", *Word Structure*, 12(1), 4-41.
- BAAYEN, H., 1992, "Quantitative aspects of morphological productivity", dans G. Booij & J. Van Marle (eds.), *Yearbook of Morphology 1991*, Dordrecht, Springer, p. 109-149.
- BAAYEN, H., 2001, *Word Frequency Distributions*, Dordrecht, Kluwer Academic Publishers.
- BAAYEN, H., 2009, "Corpus linguistics in morphology: morphological productivity", dans A. Lüdeling & M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, Berlin, Mouton de Gruyter, p. 900-919.
- BLUMENTHAL-DRAMÉ, A., 2012, *Entrenchment in Usage-Based Theories. What Corpus Data do and do not Reveal about the Mind*, Berlin, Mouton de Gruyter.

- BRINTON, L.J. & TRAUGOTT, E.C., 2005, *Lexicalization and language change*, Cambridge, Cambridge University Press.
- CARTIER, E., 2018, *Dynamique lexicale des langues : éléments théoriques, méthodes automatiques, expérimentations en français contemporain*, Habilitation à Diriger des Recherches, Université Paris 13.
- DAL, G. & NAMER, F., 2018, "Playful nonce-formations, creativity and productivity", in S. Arndt-Lappe, A. Braun, C. Moulin & E. Winter-Froemel (eds.), *Expanding the Lexicon. Linguistic Innovation, Morphological Productivity, and Ludicity*, Berlin, Mouton de Gruyter, p. 203-228.
- DOSTIE D., 2018, *Synonymie et marqueurs de haut degré. Sens conceptuel, sens associatif, polysémie*, Paris, Classiques Garnier.
- DRESSLER, W.U., BARBARESI, L.M., SCHWAIGER, S., RANSMAYR, J., SOMMER-LOLEI, S. & KORECKY-KRÖLL, K., 2019, "Rivalry and lack of blocking among Italian and German diminutives in adult and child language", dans F. Rainer, F. Gardani, W.U. Dressler & H.C. Luschützky (eds.), *Competition in Inflection and Word-Formation*, Springer, Cham, p. 123-143.
- FÁBREGAS, A., 2010, "A syntactic account of affix rivalry in Spanish nominalizations", dans A. Alexiadou & M. Rathert (eds.), *The Semantics of Nominalizations across Languages and Frameworks*, Berlin, Mouton de Gruyter, p. 67-92.
- FERRET, K., SOARE, E. & VILLOING, F., 2010, « Les noms d'événement en -age et en -ée : une différenciation fondée sur l'aspect grammatical », dans F. Neveu, V. Muni Toke, T. Klingler, J. Durand, L. Mondada, S. Prévost (éds), *Actes du congrès mondial de linguistique française 2010*, Paris, Institut de Linguistique Française, p. 945-968.
- FLAUX, N. & VAN DE VELDE, D., 2000, *Les noms en français. Esquisse de classement*, Paris, Ophrys.
- FRADIN, B., 2016, « L'interprétation des nominalisations en N-age et N-ment en français », dans E. Buchi, J.-P. Chauveau & J.-M. Pierrel, (éds), *Actes du XXVIIe congrès international de linguistique et philologie romane*, Strasbourg, Société de Linguistique Romane / Eliphi, p. 53-66.
- GARDANI, F., RAINER, F. & LUSCHÜTZKY, H.C., 2019, "Competition in Morphology: A Historical Outline", dans F. Rainer, F. Gardani, W.U. Dressler & H.C. Luschützky (eds.), *Competition in Inflection and Word-Formation*, Springer, Cham, p. 3-36.
- GREZKA, A., CARTIER, E., MATHIEU-COLAS, M., 2015, « Dictionnaires morphologiques du français contemporain : présentation de Morfetik, éléments d'un modèle pour le TAL », *Traitement Automatique des Langues Naturelles (TALN)*, Université de Caen, France.
- GUILBERT, L. & DUBOIS, J., 1961, « Formation du système préfixal intensif en français moderne et contemporain », *Le Français Moderne*, 29, p. 87-111.
- HUYGHE, R. & WAUQUIER, M., 2021, "Distributional semantics insights on agentive suffix rivalry in French", *Word Structure* 14:3.
- IZERT, M., 2011, « Archi- et ultra- : les préfixes français à valeur intensive dans leur emploi familier et dans la langue des médias », *Kwartalnik Neofilologiczny*, 58, p. 535-543.
- KOEHL, A. & LIGNON, S., 2014, "Property nouns with -ité and -itude: formal alternation and morphopragmatics or the sad-itude of the AitéN", *Morphology*, 24(4), p. 351-376.
- LANGACKER, R.W., 1987, *Foundations of Cognitive Grammar, Volume I, Theoretical Prerequisites*, Stanford, Stanford University Press.

- LIN, Y., MICHEL, J.-B., LIEBERMAN AIDEN, E., ORWANT, J., BROCKMAN, W. & PETROV, S., 2012, "Syntactic Annotations for the Google Books Ngram Corpus", in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics Volume 2: Demo Papers (ACL 2012)*.
- LINDSAY, M. & ARONOFF, M., 2013, "Natural selection in self-organizing morphological systems", dans N. Hathout, F. Montermini & J. Tseng (eds.) *Morphology in Toulouse. Selected proceedings of Décembrettes 7*, München, Lincom Europa, p.133-153.
- MANDELBROT, B., 1953, "An informational theory of the statistical structure of languages", dans W. Jackson (ed.), *Communication Theory*, London, Butterworth, p. 486-502.
- MARTIN, F., 2010, "The semantics of eventive suffixes in French", dans A. Alexiadou & M. Rathert (eds.), *The Semantics of Nominalizations across Languages and Frameworks*, Berlin, Mouton de Gruyter, p.109-139.
- MICHEL, J.B., KUI SHEN, Y., PRESSER AIDEN, A., VERES A., GRAY M., BROCKMAN, W. & the Google Books Team, 2010, "Quantitative Analysis of Culture Using Millions of Digitized Books", *Science* 331 (6014), p. 176-182.
- MISSUD, A., & VILLOING, F., 2020, "The morphology of rival -ion, -age and -ment selected verbal bases", *Lexique*, 26, p. 29-52.
- NACCARATO, C., 2019, "Agentive (para)synthetic compounds in Russian: a quantitative study of rival constructions", *Morphology*, 29(1), p. 1-30.
- REY, A. (éd.), 2016, *Dictionnaire Historique de la langue française. Nouvelle édition augmentée*, Paris, Éditions Le Robert.
- RIEGEL, M., PELLAT, J.C. & RIOUL, R., 1994, *Grammaire méthodique du français*, Paris, Presses Universitaires de France.
- SCHMID, H.-J., 2007, "Entrenchment, salience and basic levels", dans D. Geeraerts & H. Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, Oxford: Oxford University Press, p. 117-138.
- SCHMID, H.-J., 2020, *The Dynamics of the Linguistic System. Usage, Conventionalization and Entrenchment*, Oxford, Oxford University Press.
- SCHULTE, M., 2015, "Polysemy and synonymy in derivational affixation – a case study of the English suffixes -age and -ery", *Morphology*, 25(4), p. 371-390.
- THORNTON, A. M., 2012, "Reduction and maintenance of overabundance. A case study on Italian verb paradigms", *Word Structure*, 5(2), p. 183-207.
- THORNTON, A.M., 2019, "Overabundance in morphology", dans *Oxford Research Encyclopedia of Linguistics*, Oxford, Oxford University Press.
- VARVARA, R., 2018, *Verbs as nouns: empirical investigations on event-denoting nominalizations*, Doctoral dissertation, Università Degli Studi di Trento.
- ZIPF, G.K., 1965 [1935], *The Psycho-biology of Language*, Cambridge, MA: MIT Press.

NOTES

1. Nous remercions G. Dostie, S. Diwersy et le relecteur ou la relectrice anonyme pour leurs remarques et suggestions, qui nous ont permis d'améliorer la qualité de cet article.
2. Les exemples (1) sont recensés sur le site de veille néologique *Néoveille* (<http://www.neoveille.org>).

3. Les huit PHD en question sont sélectionnés à partir des listes proposées par Guilbert et Dubois (1961) et Amiot (2004), auxquelles nous avons soustrait *sur-*, dont la polyvalence d'emploi (notamment comme préposition) brouillait fortement la récolte et l'analyse des données. Nous avons en revanche ajouté *supra-*, qu'il nous semblait intéressant de comparer avec *maxi-* et *méga-*, dont l'emploi marginal ou émergent comme PHD est mentionné par Amiot (2004).

4. <https://books.google.com/ngrams>

5. Les combinaisons PHD-base citées dans les exemples (2) sont attestées dans le corpus FRCOW16A (<https://www.webcorpora.org/>) ou dans les pages francophones d'Internet (sondage effectué au moyen de Google le 03.11.20).

6. Tous les programmes et données, pour cette section et les suivantes, sont consultables à l'adresse suivante : <https://github.com/ecartierlipn/hautdegre>. Les données brutes de GNG sont accessibles à l'adresse suivante : <http://storage.googleapis.com/books/ngrams/books/datasetsv3.html>. Nous utilisons la version 2 (dite 2012) de GNG, la version 3 (comportant plus de données, et étendue à 2019) ayant été publiée après la réalisation de cette étude.

7. On appelle « *n*-gram » dans GNG une séquence de *n* (de 1 à 5) mots contigus séparés par un espace ou un trait d'union. Ces *n*-grams permettent d'une part d'identifier des collocations et certaines locutions figées et d'autre part de suivre l'évolution des usages sur la période couverte.

8. L'écart-type permet ici de mesurer la dispersion des formes lexicales selon leur fréquence d'apparition dans le corpus. L'analyse de la distribution des formes selon le rapport à l'écart-type permet d'identifier les cas les plus marginaux, dont la fréquence d'emploi est particulièrement élevée ou faible au regard de celle des autres formes.

9. Les formes analysées comme ambiguës hors contexte sont indiquées à l'aide d'une double étiquette. Ainsi *rebelle* dans *ultra-rebelle* est-il annoté par défaut comme nom et adjectif. Seules les trois combinaisons les plus fréquentes pour chaque PHD et pour le total des formes recensées sont indiquées dans le tableau 6.

10. Cette procédure ne permet cependant pas de mitiger les effets liés aux types de documents disponibles dans les différentes périodes. En particulier, les documents édités durant les deux guerres mondiales sont non seulement beaucoup moins nombreux qu'au cours des périodes contiguës, mais également de types très différents.

11. Dans la terminologie de Baayen (2009), la productivité en expansion est le ratio entre les formations nouvelles et les formations attestées. Dans le présent travail, nous approchons cette notion en indiquant dans les graphes les formations nouvelles apparues au cours d'une période et le ratio entre ce nombre et les formes attestées dans les trois périodes précédentes.

12. Techniquement, nous étudierons cette distribution de la façon suivante : afin de neutraliser certaines spécificités non pertinentes associées aux tranches annuelles, nous prendrons pour base une périodisation de dix ans, et nous comparerons, pour chaque forme construite, une période donnée à la période précédente. Si la valeur assignée à une forme à un moment donné est supérieure à 0 et que la valeur dans la période précédente était 0, alors la forme est considérée comme une apparition. Si sa valeur précédente était elle aussi supérieure à 0, alors la forme est considérée comme stabilisée. Si la valeur à un moment donné est 0 et que la valeur précédente était supérieure à 0, alors la forme est considérée comme disparue. Enfin, si la valeur est 0 dans deux périodes successives, alors la forme est catégorisée comme non active. La somme des quatre cas possibles est réalisée pour chaque période et pour l'ensemble des expressions incluant un PHD attestées dans le corpus. On peut alors tracer une courbe de l'évolution de ces cas. On trouvera dans la section 3.2. des représentations graphiques de ces évolutions.

13. Le détail des résultats présentés dans cette section et la suivante est consultable dans l'interface d'exploration des données mise en place pour l'étude : https://tal.lipn.univ-paris13.fr/neoeville/html/index_prefixes.php. Voir également note 3.

14. En dépit de la prise en compte des fréquences relatives, les deux périodes de guerre mondiale (1914-1918 et 1939-1945) induisent un biais dans l'ensemble des évolutions observées ⁽⁸⁾ peut-être dû au type de documents édités pendant ces périodes.

15. Pour suivre l'analyse présentée ici, nous invitons le lecteur à consulter et manipuler le graphe d'évolution temporelle détaillé disponible à l'adresse : https://tal.lipn.univ-paris13.fr/neoveille/html/index_prefixes.php.

16. Par convention, nous utilisons ici le signe + pour indiquer toute construction associant une base et un PHD donnés, indépendamment des variations de forme graphique observées.

17. Étant donné qu'une même base peut être exploitée par plus de deux PHD, les totaux en ligne peuvent être supérieurs à 100%. En effet, dans les situations où un PHD donné est en concurrence avec deux autres PHD ou plus pour une même base, cette base se trouve décomptée dans au moins deux combinaisons de concurrence pour le PHD en question.

RÉSUMÉS

Dans cet article, nous étudions l'évolution en français de l'emploi de huit préfixes de haut degré (PHD) entre 1800 et 2010. L'étude est fondée sur une analyse détaillée des données recueillies au moyen de *Google Ngram*. D'une manière générale, l'emploi des PHD augmente de manière significative au cours de la période, notamment entre 1900 et 1970. Un examen approfondi des fréquences d'emploi par préfixe fait apparaître des situations nuancées, avec des phases d'augmentation plus (ex. *archi-*) ou moins (ex. *super-*) précoces, un éventuel tassement en fin de période (ex. *extra-*), et une productivité plus ou moins importante selon les cas. Trois PHD (*hyper-*, *ultra-*, *super-*) dominent nettement, tant dans le nombre de formes construites que dans la fréquence d'emploi. Malgré ces disparités, la situation de surabondance affixale est avérée et persistante en diachronie. Les trois PHD les plus fréquemment employés connaissent notamment une recrudescence d'emplois en fin de période, et des cas de synonymie durables sont observés pour les formes lexicales concurrentes qu'ils permettent de construire.

This article studies the evolution between 1800 and 2010 of eight prefixes that are used to express high degree (HDP) in French. The study is based on a detailed analysis of data collected from *Google Ngram*. It appears that the use of HDPs increases considerably over the period, especially between 1900 and 1970. An in-depth examination of HDP frequencies reveals some differences, with variable productivity, earlier (e.g. *archi-*) or later (e.g. *super-*) phases of development, and possible decreases at the end of the period (e.g. *extra-*). Three HDPs (*hyper-*, *ultra-*, *super-*) clearly dominate with respect to both the quantity of output lexemes and use frequency. Despite those differences, HDP overabundance is effective and persistent in diachrony. The three most frequent HDPs are increasingly used at the end of the period, and enduring situations of synonymy are observed for competing lexical forms derived from these three prefixes.

INDEX

Mots-clés : préfixe, haut degré, concurrence affixale, surabondance, diachronie

Keywords : prefix, high degree, affix rivalry, overabundance, diachrony

AUTEURS

EMMANUEL CARTIER

Université Sorbonne Paris Nord, LIPN UMR7030 CNRS

RICHARD HUYGHE

Université de Fribourg