



HAL
open science

Pourquoi et comment utiliser la distinction écrit/oral dans l'étude des noms d'humains généraux?

Paul Cappeau, Catherine Schnedecker

► To cite this version:

Paul Cappeau, Catherine Schnedecker. Pourquoi et comment utiliser la distinction écrit/oral dans l'étude des noms d'humains généraux?. *Linguistique de l'écrit*, 2022, 3, pp.135-165. 10.19079/1de.2022.s3.5 . hal-03960270

HAL Id: hal-03960270

<https://hal.science/hal-03960270v1>

Submitted on 27 Jan 2023

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.

Pourquoi et comment utiliser la distinction écrit/oral dans l'étude des noms d'humains généraux?

Les cas de *man*, *individu* et *gens*

Paul Cappeau & Catherine Schnedecker

Présentation

- 1 L'usage oral est-il nécessaire, ou même utile, pour décrire une langue ? La réponse à cette question dépend des périodes et des traditions propres à chaque langue. Dans le domaine français, depuis le 17^{ème} siècle, la langue littéraire est considérée comme la source légitime et principale des exemples des dictionnaires (*TLF*, etc.) et des grammaires (pour la période récente à travers principalement la banque de données Frantext). Or, pour cette même langue, la pauvreté des ressources orales a été souvent pointée (Blanche-Benveniste 2000), ce qui rendait marginal tout appui sur des données suffisamment significatives. Néanmoins depuis le début des années 2000, la situation a évolué et des données orales commencent à être plus facilement accessibles : à travers des grands projets PFC, ESLO, CFPP2000, CRFP, etc. jusqu'à ORFEO (Debaisieux et Benzitoun, 2020). La question de l'intérêt de l'oral peut donc à nouveau être soulevée. Willems (1997) a fourni quelques arguments favorables tant sur le plan méthodologique que descriptif¹. Les

1 Les Anglo-Saxons ont montré tout le bénéfice qu'une description de la langue peut tirer de la diversité des situations exploitées et de la prise en compte de l'écrit ET de l'oral (Biber et *al.* 1999).

données orales conduisent à s'interroger sur la frontière entre lexicale et syntaxe (par exemple à travers l'observation des collocations), à réviser de nombreux outils descriptifs (comme la phrase ou la subordination), à identifier des zones où l'usage est flou ou hésitant ainsi qu'à formuler des hypothèses sur de possibles évolutions de la langue.

Notre apport dans cette réflexion générale sur la portée de la distinction entre oral et écrit dans la description linguistique s'appuiera sur plusieurs analyses des noms d'humains généraux (NHG) que nous avons développées. La prise en considération de l'oral nous a permis de dépasser certaines observations fondées sur les seuls exemples écrits, comme par exemple l'étude de changements linguistiques attachés aux noms dits d'humains généraux² (*homme, personne*, etc.). Cela étant, la dichotomie écrit *vs* oral ne semble pas toujours productive ou pertinente, comme nous allons tenter de le montrer à travers la présente étude.

Nous serons donc amenés dans une première partie à rappeler quelques particularités du couple oral/écrit, puis à présenter, dans une deuxième partie, les NHG et à spécifier leur intérêt pour la réflexion conduite ici. Par la suite, nous nous arrêterons sur la description de trois noms (*man, individu* et *gens*) et détaillerons les éléments d'analyse que font émerger des ressources diversifiées³. Ces données feront ressortir que la diamésie n'est qu'une des dimensions à prendre en compte et que, selon le nom examiné, des caractéristiques comme la proximité entre les locuteurs ou le genre de discours jouent un rôle majeur pour différencier les valeurs / formes décrites. Cela conduit à mettre en avant l'importance que présente la variété des données langagières pour développer des analyses pertinentes.

2 Voir *infra* partie 3 sur la pragmatization d'un mot comme *mec* et partie 5 sur la transcategorialisation de *gens*.

3 Nous avons principalement utilisé les corpus suivants : CERF (cf. note 7), MPF (cf. note 9) et Sketch Engine.

1 | Le couple oral vs écrit

- 4 Il n'est pas envisageable de faire l'inventaire de toutes les disciplines (didactique, sociolinguistique, etc.) qui contrastent oral et écrit (Goody 1993) et de préciser l'intérêt de cette distinction. On s'attachera plutôt à signaler que, de plusieurs façons différentes, cette opposition a parfois été plus ou moins remise en cause ou replacée dans un cadre plus large.

1.1. Un statut variable

- 5 Dans la tradition grammaticale et lexicographique française, si la dichotomie oral / écrit est peu présente, c'est en grande partie parce que les données orales sont très peu exploitées (Cappeau 2021). Les descriptions de la langue s'appuient quasi exclusivement sur l'usage écrit livresque (essentiellement littéraire d'ailleurs). Cette sélection des faits est toutefois partiellement masquée et la plupart des travaux affichent leur ambition de décrire « le français » (sans spécifier ou restreindre l'extension de cette désignation globale⁴). La représentation de l'oral se limite souvent à quelques exemples familiers ou à de l'oral représenté⁵ comme des répliques de théâtre ou des dialogues de romans. Trois raisons peuvent expliquer cette minoration / absence de l'oral :

- A. L'accès à des données orales. Comme rappelé *supra*, celui-ci a été longtemps entravé, réservé à des équipes restreintes, mais depuis les années 2000 et le lancement de grands projets⁶ (ESLO2,

4 On pense à des titres tels que : *Grammaire du français*, *Grammaire française*, *la grammaire pour tous*, *Grammaire méthodique du français*, *Grammaire critique du français*, *La Grande Grammaire du Français*, etc.

5 Voir Lefeuve & Parussa (éds) (2020).

6 Voir le site de l'IRCOM (<http://ircom.huma-num.fr/site/corpus.php>) pour un inventaire récent des corpus oraux.

CFPP2000, PFC, MPF...), cet obstacle a été progressivement levé. L'utilisation de données orales ne peut plus être vue comme une difficulté majeure.

- B. La forme déroutante des énoncés oraux. Une deuxième raison a pu provoquer la mise à l'écart d'exemples oraux : leur caractère souvent inhabituel ou insaisissable. En effet, par rapport à des exemples écrits publiés, l'oral non lu présente de multiples phénomènes (hésitation, bribes, répétitions, interruptions, ...) qui peuvent donner l'illusion d'énoncés désorganisés et rendre plus délicate l'analyse avec les outils classiques (tels que la notion de *phrase*, etc.). On imagine alors la réticence à intégrer, avec difficulté, de tels exemples « problématiques » aux analyses déjà construites à partir de l'écrit, sauf à être convaincu de l'apport des énoncés oraux (voir ci-après).
- C. L'apport des données orales. C'est probablement le facteur décisif même s'il est rarement mis en avant. On peut craindre que de nombreux chercheurs ne voient pas l'intérêt d'intégrer des données orales. Ils ne sont donc pas incités à élargir leurs ressources et s'appuient sur des exemples écrits (littéraire et presse), plus faciles à collecter. De plus, une longue tradition française associe fortement oral et usage familier (Blanche-Benveniste & Jeanjean 1986). La dimension normative (plus ou moins sous-jacente) des grammaires peut trouver là une raison supplémentaire d'écarter des exemples qui seraient considérés comme « déplacés » dans des ouvrages de référence.

Chez certains auteurs, la relation oral / écrit est plus vue en termes de complémentarité qu'en terme d'opposition. C'est par exemple le cas chez Damourette & Pichon (1911-1940) dont l'ouvrage comporte des exemples très variés : 6

Et c'est ici le lieu de s'inscrire formellement en faux contre les assertions des auteurs qui vont affirmant au monde que dans la

France d'aujourd'hui il y ait un écart sensible entre le langage écrit et le langage parlé.

Tout ce qui s'écrit se parle.

[...] il n'est pas une phrase écrite qui ne puisse, au moins dans certaines circonstances, venir spontanément à la bouche d'un Français de bonne compagnie. (§ 36).

- 7 C'est aussi le cas pour Blanche-Benveniste qui a constamment repoussé l'idée d'une syntaxe propre à la langue parlée, même si, dans Blanche-Benveniste (2003), elle ouvre certaines pistes de différenciation comme : l'existence de règles rédactionnelles et de filtres normatifs propres à l'écrit, « certaines tournures syntaxiques fréquentes de la langue parlée sont difficiles à "rendre" par écrit » ; de plus, elle note que « dans la langue parlée, la syntaxe est généralement très différenciée selon les « genres » de prise de parole » (p. 335).

1.2. Une variation parmi d'autres

- 8 On prendra deux exemples, éloignés dans le temps, qui permettent d'observer que de nombreux autres éléments de variation ont été repérés et signalés. L'opposition entre l'oral et l'écrit s'insère dans une série et n'est pas considérée comme plus importante ou plus significative que d'autres distinctions.
- 9 De façon peu technique mais en accordant beaucoup d'attention à des différences sensibles, Vaugelas (1647) appuie ses commentaires sur de multiples facteurs : la légitimité sociale des locuteurs (la Cour, la partie la plus saine de la Cour, les bons auteurs, les anciens écrivains, etc.), le médium, les formes du discours (la prose, les vers, la parodie, etc.) ou les styles (le style bas, etc.). La différence entre l'oral et l'écrit (qui affleure parfois au travers de formulations telles que *On dit tous les jours l'un et l'autre en parlant, mais on ne le doit point dire en écrivant que dans le style bas.* (p. 180)) n'est pas centrale.

Dans un cadre sociolinguistique, la diamésie n'est qu'une des dimensions de la variation. Ainsi, Gadet (2007) rappelle (Tableau 1) 5 types de variation parmi lesquels figure le couple oral/écrit. Cela revient à ne pas focaliser l'attention sur cette seule opposition et à manifester qu'il convient d'élargir le regard dans l'appréhension de la variation des faits langagiers :

Variation selon l'usager	temps	changement	diachronie
	espace	géographique, régional, local, spatial	diatopie
Variation selon l'usage	société, communauté	social	diastratie
	styles, niveaux, registres	situationnel, stylistique, fonctionnel	diaphasie
	chenal	oral/écrit	diamésie

Tableau 1. Représentation de la variation (repris à Gadet 2007)

1.3. Des alternatives ou remises en cause récentes

Dans plusieurs cadres, la relation de l'oral avec l'écrit a été repensée. On pense notamment à :

- la tradition allemande (Söll 1974) qui différencie médium (phonique *vs* graphique) et conception (orale ou écrite), ce qui offre quatre possibilités au lieu de deux,
- la proposition de Koch & Esterreicher (2001) qui distingue notamment le couple proximité / distance et conduit à raisonner en termes de continuum plus que de dichotomie,
- la notion de *genres* dans le cadre de la linguistique sur corpus (dans le prolongement de certaines réflexions de Bakhtine (1984), voir par exemple Biber & Conrad (2010)) qui déplace

l'opposition selon le médium vers une centration sur « la diversité des productions verbales » (Maingueneau 2004) :

Aux définitions qui laissent entendre que l'oral ne ferait pas texte, on peut objecter que, d'un point de vue textuel, il n'y a guère de différence entre un proverbe oral et un proverbe écrit, entre un slogan politique ou publicitaire écrit et un slogan politique ou publicitaire oral. Très contraints formellement, ces genres discursifs communs à l'écrit et à l'oral diffèrent par la matérialité de leur énonciation : les ressources sonores et accentuelles de l'oral et celles de la spatialité (typo)graphique du canal visuel. Ils relèvent certes de deux régimes distincts d'usage de la langue, mais ils font texte, l'un comme l'autre. (Adam 2019)

- 12 Une première série de questions émerge : dans la description de la langue, cette opposition est-elle incontournable, est-elle toujours le sésame qui éclaire les faits langagiers et comment peut-elle être exploitée en syntaxe et/ou sémantique ? L'analyse de quelques noms d'humains généraux (NHG), présentés ci-après, a pour but d'amener à réfléchir à ces questions et de s'interroger sur les éventuelles insuffisances de cette opposition oral *vs* écrit.

2 | Les noms d'humains généraux (NHG)

- 13 Ces noms (*homme, personne, gens, individu, être humain*, etc.) sont d'excellents candidats pour faire émerger un ensemble de paramètres liés à diverses formes de variation (Cappeau & Schnedecker 2014 a et b, 2015, 2017 a et b). Trois raisons principales peuvent être avancées : leur fréquence qui en permet un recueil abondant, leur sémantisme qui les prédispose au changement, leur forte sensibilité aux genres discursifs qui conduit à s'intéresser à des différences d'emploi dans les productions plus ou moins codifiées. Ce sont donc

des formes sensibles à la variation sur lesquelles on peut espérer observer l'importance des usages (ce qui ramène à la question de l'oral *vs* l'écrit).

2.1. Une fréquence élevée propice à une analyse quantitative

Les NHG sont attestés en nombre dans divers corpus (Mahlberg 2005, Benninghoven 2018) comme on va le vérifier, c'est pourquoi d'ailleurs on peut parler de noms généraux par contraste avec d'autres sous-catégories nominales (noms spécifiques, N sous-spécifiés, etc.). Plusieurs NHG sont d'une fréquence élevée aussi bien en anglais (Tableau 2) qu'en français (Tableau 3).

Noun	All Hutton	Men at work	Keyness
thing	0,03 %	0,15 %	90.1
sens	0,03 %	0,12 %	66.3
issue	0,06 %	0,19 %	63.9
people	0,08 %	0,21 %	58.6
point	0,14 %	0,29 %	45.4
story	0,07 %	0,19 %	52.9
discussion	0,05 %	0,12 %	28.0

Tableau 2. Key nouns found as subjects (Partington et al. 2013, 229)

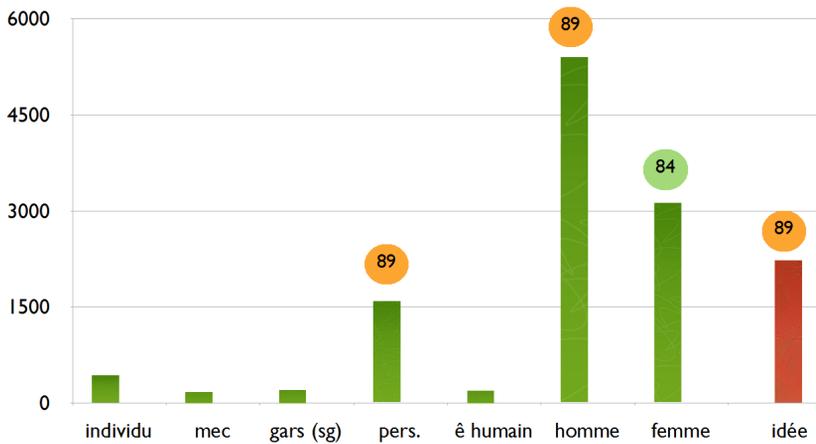
Nom	Nbre occ.	Rang	Classe de fréquence
personnes	814.158	136	7
hommes	357.280	359	8
gens	322.213	403	8
femme	306.286	435	8
femmes	300.645	441	8
homme	275.419	492	8
individus	47.369	3.102	11
individu	23.536	5.866	12

Tableau 3. Sondage de fréquence d'emploi sur un échantillon de NHG dans *Wortschatz*⁷

- 15 La conséquence de cette forte implantation dans l'usage est que ces noms peuvent être facilement récupérés et sont en nombre suffisant pour favoriser des analyses tant quantitatives que qualitatives. On a aussi pu vérifier (Cappeau & Schnedecker 2021) que certains de ces NHG présentent une répartition diversifiée et sont présents dans de nombreux corpus⁸. C'est par exemple le cas de *homme* ou *personne* (graphique 1) qui se rencontrent tous deux dans 89 des 90 sous-corpus utilisés.

⁷ https://corpora.uni-leipzig.de/fr?corpusId=fra_mixed_2012

⁸ L'observation a été conduite sur le corpus CERF-écrit (*Corpus évolutif de Référence du Français*) constitué par Jean Véronis. Ce corpus est constitué de 9 tranches "thématiques" (proche de genres) de 1 million de mots récupérés sur le web. Chaque tranche d'un million de mots (forums, presse, institutions, pratiques, etc.) est elle-même divisée en 10 sous-corpus de 100 000 mots. Par exemple, la tranche Pratiques est divisée en : horoscope, cuisine, loisirs, automobile, people, sports... C'est pourquoi le graphique 1 examine la répartition dans 90 sous-corpus.



Graphique 1. Nombre d'occurrences et répartition dans les sous-corpus du CERF

2.2. La fréquence et la généralité comme facteurs de changements

Certains NHG ont un potentiel d'évolution susceptible de les mener à diverses évolutions dont la grammaticalisation (Marchello-Nizia 2006) et la pragmaticalisation (Dostie 2004). 16

La grammaticalisation, par exemple, peut notamment faire passer un mot de la catégorie nominale au pronom. Dans de nombreuses langues, les NHG changent de catégorie (transcatégorialisation) pour se comporter comme et devenir des pronoms - on parlera de « pronominalisation » - (Haspelmath 1997, Marchello-Nizia 2006, Heine & Kuteva 2002, Bath 2002). Des pronoms indéfinis ont pour source des NHG : c'est ainsi que le pronom allemand *man* est issu du nom *Mann* et le pronom roumain *lumea* du nom *lume*. De même dans le passage du latin au français les noms *persona* et *homo* ont donné les pronoms *personne* et *on*. 17

18 La pragmaticalisation voit passer un nom au rôle de marqueur discursif (Kleinknecht 2013, Kleinknecht & Souza 2017, Alba-Juez 2009). Ce phénomène a été décrit dans de nombreuses langues et concerne, par exemple, pour l'anglais, *man*, *dude* ou *mate*, pour l'espagnol *hombre*, *mujer*, *macho*, *tío* ou *tía*, pour le swahili *bwana* (Heine & Kuteva 2002). En français, ce phénomène est illustré d'une part par *putain* (1), sorte de ponctuant du discours qui provient d'un nom d'humain désignant une profession et d'autre part par *con* (2), dont l'usage en tant que marqueur discursif est très répandu en français méridional (Pustka 2011) :

- (1) Et là je me suis dit que *putain* j'aurais pu crever quand même!
(guenillehaillon.free.fr.txt, rampé 21.10.2007)
- (2) Eh bé *con*, ma mère *con*, elle voulait m'acheter des chaussures *con*, mais *con*, elles étaient trop chères, *con* !

2.3. Variation en termes de genres discursifs

19 Nous avons déjà pu observer que les NHG tels que *gens*, *individus*, *personnes* et *hommes*, manifestent des différences d'usage particulièrement marquées. Le corpus CERF-écrit et CERF-oral (une tranche de 1 Mo de mots oraux) illustre cette inégale répartition (Tableau 4) :

	Divers	Forums	Instit.	Litt A	Polit.	Prat.	Sciences	ORAL
<i>gens</i>	414	862	213	495	119	85	62	1448
<i>individus</i>	63	73	23	6	38	8	128	15
<i>personnes</i>	195	388	882	139	477	587	275	483
<i>hommes</i>	443	395	116	524	445	139	164	224

Tableau 4. Nombre d'occurrences de certains NHG dans quelques tranches du CERF

Ces relevés font ressortir des répartitions très différentes, dans l'écrit, des NHG retenus selon les « genres » observés : par exemple, *personnes* est attesté 882 fois dans le genre dit de « textes institutionnels », et 139 fois dans les textes littéraires, soit quasiment 6 fois moins. De plus, une variation diamésique peut aussi être repérée. Le cas le plus spectaculaire est celui de *gens* (on y reviendra en partie 5) peu présent dans les écrits scientifiques, mais abondant dans les forums et plus encore à l'oral.

Pour étayer notre réflexion, nous avons choisi trois NHG qui présentent des caractéristiques très différentes : *man* est d'apparition récente en français et permet d'observer les prémices d'un fonctionnement original, *individu* est un nom humain à éclipse (il ne possède pas toujours le trait humain) ce qui le rend atypique dans notre série et il présente des connotations multiples, enfin *gens* se distingue par son importance quantitative et son caractère uniquement pluriel.

3 | Le cas de *man*

Ce nom, utilisé en français dans certains usages précisés plus loin, montre certains signes d'un processus de pragmaticalisation (cf. *supra*). Il est intéressant de regarder de plus près les données françaises pour apprécier son statut⁹.

3.1. Présentation dans les dictionnaires

Dans l'usage général, ce terme anglais est très rare en tant que mot autonome. Il est en revanche attesté dans des composés nominaux empruntés à l'anglais (*tennisman*, *rugbyman*, *one man show*, *self made man* ...) aussi bien qu'à travers des personnages de fiction – bande dessinée ou cinéma – tels *Iron man*, *Spiderman*, *Batman*, *Superman*...

9 Nous renvoyons à Cappeau & Schnedecker (2021) pour une étude comparée de *man* et *mec*.

- 24 Les dictionnaires (comme le *Grand Robert*, le *Petit Robert* ou le *TLFi*) ne comportent pas d'entrée autonome *man*, qui n'apparaît qu'au travers de la composition nominale (*self(-)made(-)man*). Absent des ouvrages de référence, ce nom est pourtant bien présent dans l'usage de certains locuteurs. D'où l'importance de constituer un corpus pour conduire (ou du moins amorcer) une analyse.

3.2. Constitution d'un corpus de travail

- 25 L'objectif principal du corpus est de disposer d'un nombre suffisant d'exemples dans lesquels *man* présente un emploi vocatif. En effet, les études précédemment citées ont montré que la pragmaticalisation est repérable, dans une première phase, à travers ces emplois. Cette hypothèse précise oriente le choix des données explorées et nous amène à privilégier les corpus récents (et non des corpus de textes littéraires plus ou moins anciens). Quatre corpus relativement récents ont été retenus (2 écrits, 2 oraux). Les résultats figurent dans le tableau 5 :

NOM ¹⁰	CERF-écrit	CERF-oral	MPF	Sketch Engine (S.E.)
Taille	9 Mo	1 Mo	750.000	1,3 milliard
Nbre occ. <i>man</i>	6 ex (4 tranches)	1 ?	6 ex	220 ex.
Médium	Écrit	Oral	Oral	Écrit

Tableau 5. Occurrences de *man* dans 4 ressources

10 CERF (*Corpus Évolutif du Français*) a déjà été présenté. MPF (*Multicultural Paris French*) est présenté dans Gadet (2017). Il s'agit d'un corpus en grande partie écologique de locuteurs jeunes de la banlieue parisienne. La ressource Sketch Engine est disponible à l'adresse suivante : <https://www.clarin.si/noske/index-en.html>.

Un important travail de « nettoyage¹¹ » a été nécessaire pour écarter des exemples non conformes comme (3) où *man-* est une amorce sans aucun lien avec le mot recherché et (4) qui contient une phrase entièrement en anglais, ce qui n'est pas le contexte souhaité. 26

(3) ils ont pas de quoi *man-* manger (CERF-oral)

(4) On retrouve cette phrase simple, "I am a man" (CERF-écrit, divers)

Une mauvaise interprétation du tableau 5 consisterait à surestimer l'opposition oral *vs* écrit ce qui nous mènerait à la conclusion (peu conforme à notre intuition et aux descriptions des travaux antérieurs) que *man* relève de l'écrit. Ce sont d'autres critères, comme le lien entre les interactants et le caractère non formel de l'interaction, les thèmes abordés voire le sexe des locuteurs, qui priment. 27

De fait, dans le CERF, le nombre d'occurrences est résiduel que ce soit à l'oral ou à l'écrit, MPF en contient légèrement plus et c'est S.E. (d'une taille imposante) qui en contient un nombre significatif. La forme *man* ne se rencontre pas dans tous les genres de textes rassemblés mais elle se concentre dans certaines productions écrites. C'est plus précisément dans les tchats et dans certains forums qu'on en recueille le plus. Une observation plus précise donne des indications supplémentaires. Les exemples dans lesquels figure *man* sont particuliers : ce n'est pas le médium qui apparaît comme déterminant mais la proximité entre les interactants que l'on repère à travers certains pronoms (*tu* dans 5), des formules de salutation (*yeah* dans 6) ou le lexique (*cool* dans 7). L'absence de *man* dans la plupart des productions (tant orales qu'écrites) du CERF tient, pour partie, à des liens plus distanciés entre des locuteurs majoritairement plus âgés. 28

11 A titre d'illustration on passe, dans CERF-écrit, de 56 exemples initiaux à 6 réellement pertinents.

- (5) Là maintenant il fait de la variété française des trucs rock'n'roll et tout (.) alors qu'avant c'était <un hip hop> man tu vois ce que je veux dire ? (Oral, MPF)
- (6) " Hell yeah, man !!! " Après six ans, ils enregistrent à nouveaux un album ensemble (Ecrit, SE)
- (7) hé les mecs je vous ai pas insulté !!!!!!! alors cool man (Ecrit, SE)

3.3. Résultats

- 29 Trop peu d'exemples se révèlent convaincants pour classer *man* comme marqueur discursif, on est donc conduit, à partir des faits recueillis, à s'intéresser aux trois positions qu'occupe ce terme dans l'énoncé. Une importante variation se dégage du tableau 6 et, sans surprise, ce sont les positions initiale et finale qui sont les plus occupées par le terme *man* :

Position initiale (8)	Position médiale (9)	Position finale (10)
77	23	120

Tableau 6. La position de *man* dans les exemples tirés de *Sketch Engine*

- (8) Eh *man* ! cela peut emmener loin si on se laisse faire la spirale va nous faire toucher le fond – (S.E.)
- (9) Sur ce coup la *man* t'as été un homme t'as ramené le croisé de Jackie Sardou (S.E.)
- (10) ... voilà pourquoi que tu vas réussir au resultat que tu souhaites allez courage *man* (S.E.)
- 30 La position médiale ou interne, qui permettrait de mieux questionner le fonctionnement de *man* en tant que marqueur discursif est

encore difficile à documenter, ce qui interdit toute conclusion définitive sur l'évolution en cours. Mais cela ne gêne en rien notre réflexion sur les données : l'utilisation du mot est récente et semble cantonnée pour l'instant à des locuteurs particuliers, dans des contextes de proximité, évoquant certains thèmes de prédilection (musique, etc.). Il semble donc probable que, pour le français, *man* en soit encore à la phase où il relève des termes « à la mode », avec une inflation d'emplois dans certains usages.

Pour poursuivre l'analyse, il faut donc envisager de rechercher 31 d'autres attestations, qui se rencontrent dans des contextes particuliers (concernant les liens entre les interactants et les thèmes notamment). En revanche, le médium (phonique ou graphique) constitue un critère moins décisif que la conception, qui doit être orale.

4 | Le cas d'*individu*

Ce NHG semble moins utilisé que certains de ses « concurrents » 32 comme *gens* ou *personnes*. Il est vrai que *individu* ne comporte le trait /+hum/ que dans une sous-partie de ses emplois (ce qui nécessite, là encore, un gros travail de tri dans les données). On ne présente que les emplois au singulier¹² (11) :

(11) Ne pensez donc pas trouver ici un quelconque plan d'entraînement standard qui pourrait s'appliquer à tout individu (tout simplement parce que ça n'existe pas !). (CERF, Pratique)

4.1. Description dans un dictionnaire

Dans la description du *Grand Robert*, le trait humain n'est pas un 33 principe de classement des valeurs mais permet de spécifier certains emplois. C'est ce que l'on a souhaité indiquer dans le tableau 7 :

12 *Individu* et *individus* présentent des propriétés distributionnelles assez différentes, ce qui explique notre choix de ne pas mélanger les deux formes.

A- Être ayant une existence propre	
1) Tout être concret	
2) En partic. Chaque être appartenant à l'espèce humaine	humain
3) Courant	
a) être, personne	humain
b) homme anonyme, indéterminé	humain
- Péj. Homme que l'on ne veut pas nommer	humain
B) Être vivant au sein d'une société	
1) Chaque unité d'une colonie, d'une société	
2) En partic. L'être humain considéré isolément dans la société	humain

Tableau 7. Les valeurs de *individu* en lien avec le trait humain (à partir du Grand Robert)

4.2. L'apport du corpus écrit

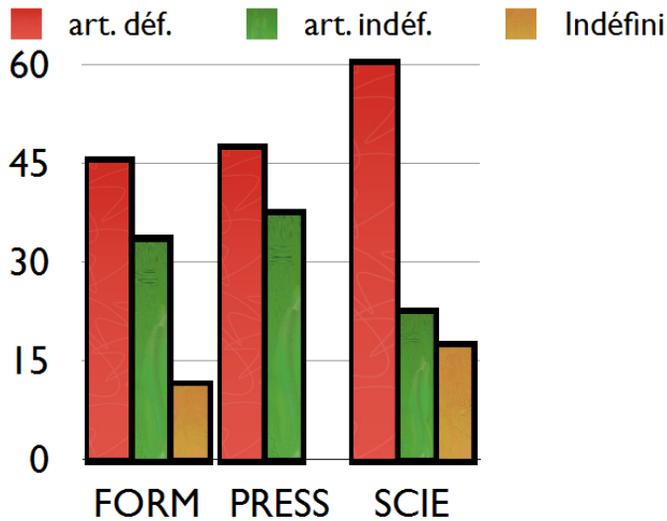
34 Dans les 9 tranches du CERF-écrit, on peut récupérer 432 occurrences de *individu* /+humain/, ce qui correspond à une moyenne de 48 exemples par million de mots. Mais une observation plus précise de la répartition (Tableau 8) témoigne de la grande hétérogénéité de la répartition selon les tranches et l'importance des thèmes ou des genres. Ainsi, *individu* est 9 fois plus attesté dans les textes scientifiques que dans les textes pratiques (horoscopes, cuisine, etc.) :

Divers	Forums	Institut.	Litt A	Litt C	Politique	Pratique	Presse	Sciences
70	91	25	24	24	31	13	30	124

Tableau 8. Répartition de *individu* /+hum/ dans les sous-corpus écrits du CERF

Plusieurs enseignements sur l'emploi de *individu* peuvent être dégagés (Cappeau & Schnedecker 2014b) :

A. La domination de l'article défini est sensible, comme on le vérifie avec le graphique 2 établi à partir des 3 tranches dans lesquelles *individu* est le plus présent.



Graphique 2. Part des articles défini, indéfini et du déterminant indéfini avec *individu*

B. Les emplois comme complément d'un nom (ou d'un adjectif) sont abondants (Tableau 9). Le nom *individu* figure donc souvent (même dans les forums) dans des configurations syntaxiques illustrées en (12) et (13) :

- (12) La survie de l'individu dépend de son combat contre cette déshumanisation (CERF, Forums)

- (13) qui permette l'émancipation de l'individu dans sa vie intime et dans sa vie sociale (CERF, Politique)

Forums	Presse	Sciences
30 %	43 %	39 %

Tableau 9. Proportion d'emplois de *individu* comme complément de nom (ou d'adjectif)

C. Peu d'adjectifs sont associés au nom *individu*, et ceux qui le sont sont plutôt spécifiques, liés aux thèmes du sous-corpus. A titre d'exemple, on peut citer *quelconque* (14) en philosophie, *extraordinaire* (15) en littérature et *isolé* (16) en sciences. A ce point de vue, *individu* manifeste un comportement similaire à celui des « shell nouns » (Schmid 2000) comme *fact* dont le sémantisme vague est suppléé par des constructions en *que* (*le fait que...*) ou des adjectifs particuliers.

- (14) Car le philosophe n'est pas un *individu quelconque* "parlant" en tant que philosophe (CERF, Divers)
- (15) C'est un *individu extraordinaire* et, cependant, il ne présente rien de réellement insolite (CERF, Litt A)
- (16) aucun être humain ne peut être considéré comme un *individu isolé*. Il est membre d'une société (CERF, Sciences)

36 Cette particularité met en lumière l'importance du « genre ». On peut aussi signaler les faits divers dans les genres de la presse dans lesquels *individu* renvoie assez systématiquement aux malfaiteurs et prend donc une coloration souvent négative notamment par le biais d'adjectifs (*individu masqué, cagoulé, armé*).

4.3. L'apport du corpus oral

En quoi l'appui sur des données orales modifie-t-il ou améliore-t-il les analyses précédentes ? Du point de vue quantitatif, ce nom est faiblement présent dans plusieurs corpus que nous avons consultés. On dénombre :

- 15 occurrences dans CERF-oral (1 million de mots)
- 7 dans MPF (750 000 mots)
- 3 dans des interviews politiques (300 000 mots)

Cette rareté fragilise fortement la possibilité de conduire une analyse détaillée. Pour ce qui concerne les propriétés distributionnelles, on retrouve les tendances signalées précédemment : domination de l'article défini, emploi fréquent comme complément d'un nom (environ 30 % des cas) et absence d'adjectif¹³. Aucune information nouvelle n'y apparaît et rien ne contredit les indications qu'a apportées l'analyse de l'écrit. L'oral n'apporte finalement aucun éclairage original. De plus, le faible nombre d'exemples attestés n'aurait pas permis de proposer une classification robuste des faits.

On peut toutefois considérer que l'opposition oral *vs* écrit présente un intérêt dans la description de *individu /+hum/* : c'est un nom qui se rencontre nettement plus à l'écrit et occupe une position marginale dans les productions orales. Mais la description doit plutôt s'appuyer sur les genres pour faire ressortir des particularités d'emplois. Donc plus qu'au médium, c'est à cet autre facteur qu'il convient d'accorder une place centrale.

13 Cette dernière observation fait bien écho à l'ajout rare d'adjectifs à l'écrit. On sait qu'à l'oral les adjectifs sont moins utilisés.

5 | L'exemple de *gens*

40 Cette forme présente de nombreux intérêts pour l'analyse linguistique et en particulier conduit à s'interroger sur la possible grammaticalisation du nom *gens*. En effet, à l'instar de plusieurs langues, le français a déjà réalisé ce type d'évolution avec *personne*. Et il n'est pas exclu que *gens* suive la même voie. On précisera quelles propriétés appuient cette hypothèse et quelles données facilitent les observations les plus intéressantes.

5.1. Description dans un dictionnaire

41 La définition du *TLFi* comporte quatre sous-parties rapidement indiquées ci-après :

A. personnes (en nombre indéterminé).

C'est la partie la plus développée où l'on signale des quantifieurs (*peu de...*), des adjectifs (*austères...*), certaines associations (*braves gens, bonnes gens...*), des compléments de noms (*d'esprit, de goût...*), des locutions (*gens du monde, jeunes gens...*).

B. *gens de...* + nom de profession

C. adjectif possessif + *gens* (vieilli)

D. *droit des gens*. Voir *gent*

42 Dans le *Grand Robert*, une remarque initiale sur la quantification est fournie : jusqu'au 18^{ème} siècle, les numéraux étaient compatibles avec *gens* (*dix gens*). Actuellement, seuls les approximatifs sont utilisés (*une centaine, une douzaine de gens*). Cet emploi est encore possible lorsque le nom *gens* est modifié par des adjectifs comme *pauvres, petites*, etc. (*trois pauvres gens*).

43 Cela étant dit, ces ouvrages de référence ne font aucune allusion à un possible changement catégoriel et ne donnent pas d'indication

sur la densité d'emplois (ou les variations de densité selon les productions).

5.2. L'apport du corpus écrit

Le corpus CERF confirme l'importance quantitative de *gens* et fait aussi ressortir une forte variation selon les tranches du corpus (Tableau 10) :

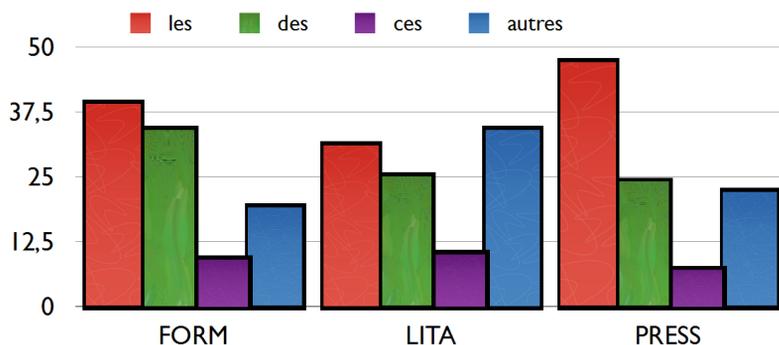
Divers	Forums	Institut.	Litt A	Litt C	Politique	Pratique	Presse	Sciences
415	849	213	486	339	117	87	315	59

Tableau 10. Répartition de *gens* dans les sous-corpus écrits du CERF

Le nom *gens* est moins utilisé dans les corpus de langue « technique » (sciences, pratiques et institutionnels) que dans le corpus forums. A ce stade, on peut se demander si la variation est liée aux genres / thèmes abordés dans les corpus (comme on vient de le voir pour *individu*) ou si un autre facteur pourrait jouer. On pense en particulier au trait /familier/ qui a longtemps été accolé à *gens* (le *Petit Robert* de 2008 en faisait encore état).

Plusieurs propriétés distributionnelles peuvent être dégagées à partir des données :

- A. La plus ou moins grande domination de l'article défini. Le graphique 3 montre à la fois la diversité des déterminants qui précèdent le nom *gens* et la spécificité du corpus de Presse où l'article défini est plus nettement dominant :



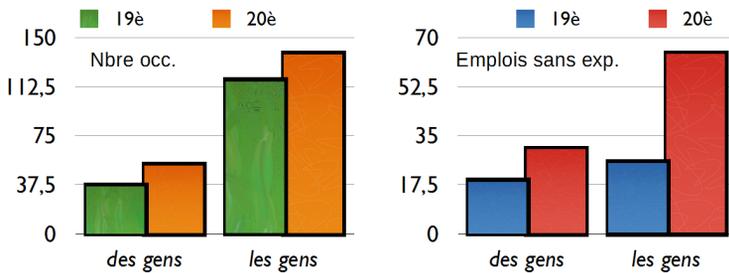
Graphique 3. Répartition des voisins de gauche de gens dans 3 sous-corpus du CERF

- B. Pour ce qui concerne les adjectifs, deux groupes peuvent être établis : d'une part, des adjectifs que l'on rencontre dans plusieurs tranches du CERF, tels que *jeunes*, *braves*, *petites*, *pauvres*, *bonnes*, *honnêtes*, et qui entrent souvent dans des collocations tirant vers la polylexie (*jeunes gens*, *petites gens*) ; d'autre part, des adjectifs dont l'apparition est limitée à un sous-corpus (*vraies*, *heureuses* dans forums *vs vieilles* ou *viles* dans la littérature).
- C. Le nom *gens* est assez fréquemment expansé. C'est en littérature que cette configuration est la plus observée. Les prolongements sont de deux types, comme indiqué dans le tableau 11 :

	Forums	Litt A	Presse
Total	271 (31 %)	207 (42 %)	67 (21 %)
Complément de N	88	109	24
Relative ou <i>que</i> forme	183	98	43

Tableau 11. Expansions à droite de gens dans 3 sous-corpus du CERF

D. En élargissant un peu l'empan d'observation, on peut signaler une évolution des emplois de *gens* en diachronie, en littérature¹⁴. Les deux faits d'évolution les plus significatifs, entre le 19^{ème} et le 20^{ème} siècles, concernent : une augmentation sensible du nombre d'occurrences (Graphique 4) et une croissance importante des emplois sans expansion (Graphique 5) :



Graphique 4. Le nombre d'occurrences Graphique 5. Les emplois sans expansion

5.3. L'apport du corpus oral

Dans la tranche orale du corpus CERF¹⁵, *gens* manifeste un comportement différent, sous quatre angles, de ce qui a été observé précédemment : 47

- A. Du point de vue quantitatif, le nombre d'occurrences explose dans le corpus oral (1431 exemples soit 65 % de plus que la tranche écrite la mieux dotée).
- B. Un double mouvement est perceptible en ce qui concerne la répartition des déterminants : d'une part, le total des occurrences

14 On a utilisé pour cette analyse (Cappeau & Schnedecker 2014a) deux corpus de 600000 mots : l'un de littérature du 19^{ème}, l'autre de romans du 20^{ème}.

15 Les enregistrements oraux du CERF datent de la période 1985-2000.

du mot *gens* déterminé par l'article défini (*les*) et indéfini (*des*) progresse fortement (85% contre 73% dans le corpus Forums), ce qui tend à marginaliser l'emploi d'autres déterminants. D'autre part, la régression du défini (aux dépens de l'indéfini) s'accroît (*les* constitue plus de la moitié des emplois du déterminant).

C. Les emplois non expansés de *gens* progressent aussi (73% dans le corpus oral contre 64 % dans Forums). Cette propriété peut être corrélée à deux facteurs :

- les adjectifs sont peu nombreux et peu variés. Il s'agit principalement de locutions ou de polylexies (*jeunes, pauvres, braves et honnêtes*),
- parmi les voisins de droite, on rencontre fréquemment *qui* (dans 37 % des expansions), unité qui n'introduit pas nécessairement une relative mais est plus en lien avec une tournure de type *il y a... qui* (17).

(17) il y a des gens qui parleront très bien qui parleront facilement (CERF, Oral)

D. Enfin une autre différence peut être observée à l'oral en ce qui concerne la fonction syntaxique d'un syntagme avec *gens* non expansé : 42% des occurrences de *les gens* et 18% de *des gens* remplissent la fonction sujet (18) et (19). Cela est d'autant plus singulier que les sujets, à l'oral, sont plus souvent réalisés sous forme de pronoms que de SN (Blanche-Benveniste 1994, Lambrecht 1987).

(18) puis les gens ont pas été informés (CERF, Oral)

(19) vous voyez que + des gens défilent et protestent (CERF, Oral)

L'ensemble de ces changements nous a conduits à formuler 48 l'hypothèse d'un changement catégoriel de *gens*. Actuellement, deux formes cohabiteraient : un *gens* nominal dont la syntaxe est libre et un *gens* pronominal, d'apparition récente, plus nettement repérable à l'oral, dont la syntaxe est réduite ou contrainte. Le tableau 12 résume les propriétés comparées de ces deux formes :

<i>gens nominal</i> (20)	<i>gens pronominal</i> (21)
variété des déterminants	déterminant <i>les</i> ou <i>des</i>
variété des expansions	pas d'expansion
toutes fonctions syntaxiques	spécialisé en fonction sujet
divers synonymes en contexte	équivalent de <i>on</i> , <i>les hommes</i>

Tableau 12. Comparaison des propriétés des deux *gens*

(20) nous recrutons des agents des gens qui ont une petite + une formation agricole (CERF, oral)

(21) les gens ils s'imaginent que nous on peut avoir nous on (n') a pas le droit en fait d'avoir nos sautes d'humeur (CERF, oral)

L'hypothèse de l'évolution, actuellement en cours, de *gens* a été 49 corroborée par l'observation du corpus MPF (plus récent et issu de jeunes locuteurs) qui accentue les tendances décrites à partir de CERF-oral (Cappeau & Schnedecker 2018).

Pour résumer, la forme *gens* présente des changements 50 importants :

- sous l'angle diachronique, entre un écrit antérieur (19^{ème}) et un écrit récent (20^{ème})¹⁶,

16 Les différences sont encore plus frappantes quand on s'appuie sur un corpus littéraire du 17^{ème} (Cappeau & Schnedecker 2015).

- dans la période contemporaine, entre des données écrites et des données orales,
 - à l'oral, entre le CERF (plus ancien et composé de locuteurs plus âgés) et MPE.
- 51 Pour fournir une analyse détaillée de cette forme *gens*, la variation doit alors être saisie sur deux plans :
- en diachronie, pour ce qui concerne l'écrit, mais aussi pour l'oral (diachronie courte),
 - en diamésie, où l'oral permet de défendre la piste d'une accélération en cours de changement et d'une évolution catégorielle.

6 | Conclusion

- 52 On espère avoir montré, à travers les trois études de cas retenues, que les NHG se prêtent bien à une discussion sur l'importance et la pertinence de l'opposition oral *vs* écrit. Cette distinction présente un réel intérêt – et l'on peut regretter que, parfois, l'oral soit délaissé – à travers la diversification des sources et des locuteurs analysés (*i.e.* pas seulement des usagers experts de l'écrit publié). Inclure l'oral donne accès à un matériau linguistique plus diversifié, moins soumis à la norme ou aux contraintes (routines) d'écriture. L'oral permet aussi de disposer d'usages « avancés » difficiles à documenter.
- 53 Toutefois, s'en tenir à l'opposition de médium fait courir le risque de négliger d'autres paramètres qui peuvent compléter et enrichir les analyses. On rappellera les autres aspects que les NHG étudiés ont permis de mettre en lumière :
- la proximité *vs* la distance entre les intervenants (cas de *man*) et la conception du message plus que le médium utilisé,

- les genres ou domaines discursifs qui font peser des contraintes d'emplois (cas de *individu*).

Ce travail montre aussi la difficulté à travailler sur un corpus unique 54 car divers paramètres peuvent entrer en jeu et imposent de diversifier les ressources pour disposer des ressources pertinentes en vue des analyses.

Références

- ADAM Jean-Michel (2019) “La notion de texte”, In: *Encyclopédie grammaticale du français*, 1–75.
- ALBA-JUEZ Laura (2009) “Little words in small talk: some considerations on the use of the pragmatic markers man in English and macho/tio in Peninsular Spanish”, In: R.P. Leow, H. Campos & D. Lardiere (eds.), *Little words: their history, phonology, syntax, semantics, pragmatics and acquisition*, Washington, D.C., Georgetown University Press, 171–181.
- BACHTIN Michail (1984) *Esthétique de la création verbale*, Paris, Gallimard.
- BATH Darbhe Narayana Shankara (2002) *Pronouns*, Oxford, Oxford University Press.
- BENNINGHOVEN Vera (2018) *The functions of ‘general nouns’: theory and corpus analysis*, Bern, Peter Lang.
- BIBER Douglas, CONRAD Susan (2019) *Register, genre, and style*, Cambridge, Cambridge University Press.
- BIBER Douglas, JOHANSSON Stig, LEECH Geoffrey, CONRAD Susan, FINEGAN Edward (1999) *Longman grammar of spoken and written English*, London, Longman.
- BLANCHE-BENVENISTE Claire (1994) “Quelques caractéristiques grammaticales des sujets employés dans le français parlé des conversations”, In: M. Yaguello (éd), *Subjecthood and subjectivity: the status of the subject in linguistic theory*, Paris, Ophrys, 77–108.
- (2003) “La langue parlée”, In: M. Yaguello (éd), *Le grand livre de la langue française*, Paris, Seuil, 317–344.
- BLANCHE-BENVENISTE Claire, JEANJEAN Colette (1987) *Le français parlé: Éditions et transcription*, Paris, Didier.
- BLANCHE-BENVENISTE Claire, WILLEMS Dominique, MERTENS Piet (eds) (1990) *Le français parlé: études grammaticales*, Paris, CNRS.
- CAPPEAU Paul (2021) *Une grammaire à l'aune de l'oral?*, Rennes, Presses universitaires de Rennes.
- CAPPEAU Paul, SCHNEDECKER Catherine (2014a) “Des gens bien différents à l'écrit et à l'oral: Hypothèses sur la pronominalisation des SN en gens”, *Verbum* 36, 55–74.
- (2014b) “Gens, personne(s), individu(s): Trois saisies de l'humain”, *SHS Web of Conferences* 8, 3027–3040.

- (2015) “(Les/des) gens vs (les/des) personnes: évolution diachronique et comparaison oral / écrit. Des SN en voie de pronominalisation ?”, In: K. Jeppesen Kragh & J. Lindschouw (éds), *Les variations diasystématiques et leurs interdépendances dans les langues romanes*, Strasbourg, Éditions de linguistique et de philologie, 449–462.
- (2017a) “L’oral des jeunes fait-il évoluer la langue ?”, In: F. Gadet (éd), *Les parlers jeunes dans l’Ile-de-France multiculturelle*, Paris, Ophrys, 127–142.
- (2017b) “A quoi tiennent les changements des noms d’humains généraux ?”, *Langages* 208, 95–112.
- (2021) “Les noms d’humains généraux: contribution à la différenciation noms sous spécifiés/noms généraux”, *Corela* 34 (HS).
- DAMOURETTE Jacques, PICHON Édouard (1911-1940) *Essai de Grammaire de la Langue Française*, Paris, d’Artrey.
- DEBAISIEUX Jeanne-Marie, BENZITOUN Christophe (2020) “Présentation [d’Orféo]”, *Langages* 219, 9–24.
- DOSTIE Gaétane (2004) *Pragmaticalisation et marqueurs discursifs: Analyse sémantique et traitement lexicographique*, Bruxelles, De Boeck.
- GADET Françoise (2007) *La variation sociale en français*, Paris, Ophrys.
- GADET Françoise (ed) (2017) *Les parlers jeunes dans l’Ile-de-France multiculturelle*, Paris, Ophrys.
- GOODY Jack (1993) *Entre l’oralité et l’écriture*, Paris, Presses Universitaires de France.
- HASPELMATH Martin (1997) *Indefinite pronouns*, Oxford, Clarendon Press.
- HEINE Bernd, KUTEVA Tania (2002) *World lexicon of grammaticalization*, Cambridge, Cambridge University Press.
- JEPPESEN KRAGH Kirsten, LINDSCHOUW Jan (eds) (2015) *Les variations diasystématiques et leurs interdépendances dans les langues romanes*, Strasbourg, Éditions de linguistique et de philologie.
- KLEINKNECHT Friederike (2013) “Mexican güey: from vocative to discourse marker. A case of grammaticalization ?”, In: B. Sonnenhauser & P. Noel Aziz Hanna (eds.), *Vocative!: addressing between system and performance*, Berlin, de Gruyter, 235–268.

- KLEINKNECHT Friederike, SOUZA Miguel (2017) "Vocatives as a source category for pragmatic markers: from deixis to discourse marking via affectivity", In: C. Fedriani & A. Sansó (eds.), *Pragmatic markers, discourse markers and modal particles: new perspectives*, Amsterdam, Benjamins, 257–287.
- KOCH Peter, OESTERREICHER Wulf (2001) "Langage parlé et langage écrit", In: G. Holtus, M. Metzeltin & C. Schmitt (éds), *Lexikon der romanistischen Linguistik (LRL): I/2*, Tübingen, Niemeyer, 584–627.
- LAMBRECHT Knud (1987) *Topic, focus and the grammar of spoken French*, Ann Arbor, University of Michigan.
- LEFEUVRE Florence, PARUSSA Gabriella (2020) "L'oral représenté en diachronie et en synchronie: une voie d'accès à l'oral spontané ?", *Langages* 217, 9–21.
- LEOW Ronald P., CAMPOS Héctor, LARDIERE Donna (eds) (2009) *Little words: their history, phonology, syntax, semantics, pragmatics and acquisition*, Washington, D.C., Georgetown University Press
- MAHLBERG Michaela (2005) *English general nouns: a corpus theoretical approach*, Amsterdam, Benjamins.
- MAINGUENEAU Dominique (2004) "Retour sur une catégorie: le genre", In: J. Adam, J. Grize & M. Ali Bouacha (éds), *Texte et discours: catégories pour l'analyse*, Dijon, Éditions universitaires de Dijon, 107–118.
- MARCHELLO-NIZIA Christiane (2006) *Grammaticalisation et changement linguistique*, Bruxelles, De Boeck.
- PARTINGTON Alan, DUGUID Alison, TAYLOR Charlotte (2013) *Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies (CADS)*, Amsterdam, Benjamins.
- PUSTKA Elissa (2011) "L'accent méridional: représentations, attitudes et perceptions toulousaines et parisiennes", *Lengas* 69, 117–152.
- SÖLL Ludwig (1974) *Gesprochenes und geschriebenes Französisch*, Berlin, E. Schmidt.
- VAUGELAS Claude Favre (1970) *Remarques sur la langue française*, Genève, Slatkine.
- WILLEMS Dominique (1997) "Histoire, linguistique et sources orales", *Recherches sur le français parlé* 14, 11–19.

