

Contribution empirique à l'étude du gérondif et du participe présent en français parlé et écrit¹

Marie-Pierre Escoubas-Benveniste¹, Oreste Floquet², Sergio Bolasco³

¹ MEMOTEF, Sapienza, Università di Roma – mariepierre.escoubas@uniroma1.it

² DSEAI, Sapienza, Università di Roma – oreste.floquet@uniroma1.it

³ MEMOTEF, Sapienza, Università di Roma – sergio.bolasco@uniroma1.it

Abstract

The purpose of our contribution is to study the distribution and use of present participle (i.e. *sortant*) and french “gérondifs” (i.e. *en sortant*) in spoken and written French, which is a question that has not yet been discussed in the literature, as far as we know. By oral French we mean spontaneous oral productions by native speakers including unformal and unguided conversations as well as guided oral interactions; by written French, the formal written language used in contemporary news papers. Empirical data analysis shows evidence that frequency of use and meaning effects of both forms considerably vary according to the type of corpus: both verbal modes are rare in oral speech, they are concentrated on very few verbs and grammaticalized forms. More over their distribution over spoken and written language show a reverse tendency, “gérondifs” being more frequently used than present participles in spoken French whereas present participles are more numerous than *gérondifs* in written French. This fact could explain why “gérondif” forms may take over some values usually attributed to present participles in written French. The data used for the linguistic analysis – verbal forms and their concordances extracted from corpora of oral and written French– are the results of a semi-automatic computer assisted analysis via TaLTaC2, performed on non lemmatized corpora following a procedure of graphical forms extraction through local grammar based queries, categorization, lexicalization of the verbal forms and retrieval of the concordances.

Résumé

Notre recherche est centrée sur la question, peu étudiée, de la présence effective du gérondif (i.e. *en sortant*) et du participe présent (i.e. *sortant*) en français parlé et écrit. Par français parlé, nous entendons les productions orales spontanées – entretiens libres ou guidés – de locuteurs dont la langue maternelle (ou une des langues maternelles) est le français ; par français écrit, la langue de la presse française contemporaine. L'analyse des données empiriques montre que la fréquence d'usage et les valeurs sémantiques des deux formes varient beaucoup selon le type de corpus examiné : ces deux modes verbaux sont peu utilisés en français parlé, ils concernent un nombre très limité de verbes et apparaissent souvent dans des structures grammaticalisées. De plus, leur distribution à l'oral et à l'écrit suit une tendance inverse, le gérondif étant d'un usage plus fréquent que le participe présent en français oral alors que le participe présent est plus fréquent que le gérondif en français écrit. Ce phénomène pourrait s'expliquer par la tendance qu'ont les formes gérondives à l'oral à prendre en charge certaines valeurs d'ordinaire attribuées aux participes présents en français écrit. Le traitement des données utilisées pour l'analyse linguistique – formes verbales et leurs concordances extraites de corpus de français parlé et écrit – est le résultat d'une analyse textuelle

1 Cette étude est le fruit d'une collaboration entre les trois auteurs tant sur le plan théorique que sur le plan pratique ; néanmoins Marie-Pierre Escoubas-Benveniste a écrit les chapitres 1, 4 et 5, Oreste Floquet le chapitre 2, Sergio Bolasco le chapitre 3.

semi-automatique au moyen du logiciel Taltac2 effectuée sur des corpus non lemmatisés selon la procédure suivante: extraction des formes graphiques par requêtes élaborées au moyen de grammaires locales, catégorisation, lexicalisation des segments répétés et extraction des concordances.

Mots-clés : français parlé, français écrit, gérondif, participe présent, linguistique de corpus, traitement automatique des langues.

1. Introduction

À la différence d'autres langues romanes et en particulier de l'italien, le français possède deux expressions morphologiques distinctes pour le verbe propre et pour le verbe impropre : le gérondif et le participe présent². Si le problème « ontologique » concernant la morphologie du gérondif et du participe présent et leur degré de ressemblance mutuelle a fait l'objet de recherches sur ces deux formes non finies du verbe, on peut dire en revanche, que le terrain des analyses de leurs usages concrets reste à défricher presque complètement³. Rappelons brièvement qu'il y a au moins deux positions opposées : (a) le gérondif n'est qu'un participe avec préposition, (b) le gérondif et le participe présent sont deux modes distincts à part entière. Or, le point focal de notre intervention consiste non pas à évacuer le problème, mais plutôt à apporter de nouvelles données empiriques pouvant sinon le trancher au moins l'éclairer sous un nouveau jour. Il s'agira moins de fournir d'autres arguments en faveur de l'une ou de l'autre position que de démontrer qu'une analyse des usages réels de ces deux modes verbaux intégrant à la fois les aspects quantitatifs et qualitatifs enrichit considérablement et la description et l'interprétation.

Cette intervention, par ailleurs, représente le deuxième volet d'une plus ample enquête que nous sommes en train de mener à propos de l'incidence statistique des formes non finies du verbe en fonction des situations discursives parlée et écrite. Dans une précédente étude (Floquet *et al.*, 2012) nous avons étudié, pour l'essentiel, l'usage différentiel du gérondif selon la modalité discursive. Tout d'abord, nous avons essayé de montrer que sa distribution est quantitativement si peu homogène qu'on a souvent l'impression que, dans le discours oral spontané, il occupe une position marginale par rapport à l'écrit. En outre, du point de vue lexical, nous avons constaté, après avoir décortiqué minutieusement toutes ses occurrences, que le gérondif peut appartenir à des formes grammaticalisées déjà répertoriées (du type *en supposant que*) et à d'autres (i.e. *en parlant de*) non encore décrites comme telles. Enfin, du point de vue distributionnel, l'analyse des concordances du gérondif à l'oral semble soutenir l'hypothèse que le gérondif puisse prendre en charge certaines valeurs exprimées à l'écrit par le participe présent.

L'objectif de cette étude est donc de tenter de fournir la pièce manquante à la description que nous avons entreprise de ces verbes du français en intégrant à l'analyse différentielle des usages oraux et écrits du gérondif celle du participe présent.

2 Voir sur cet aspect (Floquet, 2011). Les verbes sont des formes verbales non finies qui nécessitent, pour être interprétées sémantiquement la présence d'un co-texte et d'un contexte. Ils sont dits « propres » s'ils représentent une modification circonstancielle de la prédication primaire et « impropres » s'ils correspondent à une proposition secondaire subordonnée à la principale.

3 Un premier recensement des usages se trouve dans Rossi-Gensane (2011 : 88) : « On remarquera enfin que l'oral a abondamment recours aux verbes, mais beaucoup moins aux participes. [...] Les participes présents (par exemple du type *Mangeant peu, il ne grossissait pas*) y sont en général peu représentés » ; (gras et italique de l'auteur).

2. Les corpus de français oral spontané et de français écrit

2.1. Corpus d'oral (PFC et CFPP)

Pour l'étude de l'oral spontané⁴, nous avons eu recours à deux corpus distincts. En effet, par rapport à notre étude séminale qui était limitée pour l'oral au corpus Phonologie du français contemporain (dorénavant PFC), nous avons souhaité sinon fournir un échantillon quantitativement équivalent à celui de l'écrit, tout au moins nous donner la possibilité de vérifier nos observations initiales sur un corpus comparable (volume et type d'oral spontané). C'est pourquoi nous avons ajouté au corpus PFC le corpus du Français parlé parisien (dorénavant CFPP).

Le projet international PFC (Phonologie du Français Contemporain) est l'un des plus importants corpus de français parlé qui sont actuellement disponibles pour mener des recherches phonologiques, morpho-syntaxiques et discursives⁵. Chaque enquête suit un protocole commun et est à son tour divisée en quatre parties : une discussion libre, une discussion guidée, une lecture de texte et de liste de mots (toujours les mêmes). Intéressés que nous sommes par l'incidence des syntagmes gérondifs dans le français tel qu'il est parlé spontanément, nous avons choisi d'analyser la discussion libre et la discussion guidée. Par souci d'homogénéité avec la variété diatopique de français du corpus écrit (presse française) nous nous sommes limités aux enquêtes menées sur le territoire métropolitain (15 départements répartis sur tout le territoire, pour un total de 280 dialogues analysés ; la partie de PFC représente 529 543 occurrences.

Le Corpus de Français Parlé Parisien (CFPP) est, en revanche, composé d'un ensemble d'interviews portant sur la ville : les quartiers de Paris et de la proche banlieue. Les sujets sont tous des habitants de la région Ile-de-France, cependant compte tenu des caractéristiques démographiques de la population de cette région, les auteurs ont jugé utile de sélectionner les personnes selon des critères extralinguistiques explicites afin de constituer, pour chaque département, un groupe de référence représenté par des locuteurs français natifs. Le corpus est constitué d'enregistrements longs (entre 60 et 105 minutes, soit 20 000 mots en moyenne), ce qui permet par exemple de caractériser des idiolectes. Il correspond à 36,6 heures d'enregistrement soit 483 152 occurrences⁶.

4 Quand nous parlons d'oral spontané, nous entendons les manifestations orales s'opposant à toutes ces formes d'oral formel (radio, télévision, conférence, etc.) qui sont le plus souvent de l'écrit oralisé. Pour ce qui est du protocole adopté dans toutes les enquêtes du PFC, nous rappelons brièvement que ce qui différencie la discussion libre et la discussion guidée c'est que dans le premier cas, il n'y a aucune spécialisation des rôles, l'enquêteur participant, lui-même, à la conversation. La discussion guidée, en revanche, met face à face un enquêteur et un enquêté. Soulignons aussi que dans le choix des informateurs, on suit toujours la méthode dite de la proximité sociale et de la familiarité. Les interviewés sont des personnes déjà connues de l'enquêteur (amis, parents, etc.), si bien que l'échange qui en résulte est moins formel et plus spontané que s'il s'agissait de personnes qui n'entretiennent aucun rapport entre elles.

5 .Le corpus est disponible à l'adresse www.projet-pfc.net

6 Le corpus est disponible à l'adresse <http://cfpp2000.univ-paris3.fr/>

2.2. *Corpus de français écrit*

Pour l'analyse du français écrit, nous avons choisi d'explorer un corpus de presse, mis à notre disposition par Jean-Marie Viprey qui regroupe les années 2006-2010 du *Monde*, du *Figaro*, de *Libération*, de *L'Humanité* et de *La Croix*. Le corpus de presse est une ressource qui appartient au «Scriptorium numérique FANUM» de l'Université de Franche-Comté, d'un volume de 2,2 GB et de plus de 400 millions d'occurrences. Nous nous sommes limités à l'analyse d'une partie du corpus, constitué par l'ensemble des articles de *L'Humanité* de l'année 2010 (ci-après LH2010) qui représente 40 MB de texte équivalant à 6 873 329 occurrences, dont la taille est suffisante pour garantir la stabilité des fréquences.

3. Méthode d'exploration des corpus et de construction des données

3.1. *Exploration assistée par le logiciel Taltac*

Une première recherche effectuée sur une année de parution de chacun des quotidiens a montré que les formes en /-ant/, constitutives du gérondif et du participe présent, sont quantitativement stables indépendamment du journal considéré. On peut donc faire l'hypothèse que les journaux s'équivalent en ce qui concerne l'usage du gérondif et du participe présent. Par conséquent, nous avons limité l'analyse à l'année 2010 de *L'Humanité*, ce qui nous permet de compléter les mesures quantitatives par une analyse linguistique plus fine effectuée sur un corpus de concordances de taille raisonnable.

Le repérage des données de l'analyse linguistique – formes verbales et contexte d'apparition – est le résultat d'une analyse semi-automatique assistée au moyen du logiciel TaLTaC2 (Bolasco, 2010) qui opère sur le corpus LH2010 et sur les transcriptions écrites de l'oral spontané (PFC⁷, CFPP⁸), tous deux non lemmatisés. Le logiciel permet, entre autres fonctions, l'extraction de formes graphiques, la catégorisation et la lexicalisation des séquences extraites répertoriées comme des formes graphiques complexes (*multiword expressions*) ainsi que l'extraction des concordances de ces séquences lexicalisées.

3.2. *Procédure d'extraction des occurrences du gérondif*

Le gérondif en français étant une forme discontinue qui autorise l'insertion d'éléments entre /en/ et la forme en /-ant/ du verbe, il faut dresser la liste des éléments susceptibles d'être insérés afin de pouvoir effectuer une requête complexe basée sur la grammaire locale de cette forme verbale. Nous recherchons donc les segments bornés à gauche par /en/, à droite par /-ant/ et permettant l'insertion de 0, 1, 2, ou 3 éléments syntaxiques parmi les 18 recensés (pronoms clitiques et négation ainsi que leur forme élidée)⁹. Une requête par expression régulière permet, après avoir catégorisé les 18 candidats à l'insertion (CATSEM(insert) selon

7 Le protocole PFC prévoit une transcription orthographique standard sans normalisation des éléments morphémiques pouvant être absents (le type *y a pas* pour *il n'y a pas*). La ponctuation reflète, par ailleurs, l'interprétation syntaxique du transcripateur.

8 Le protocole adopté pour le CFPP prévoit une transcription orthographique standard sans ponctuation, ni points, ni virgules, mais des indications des pauses et des allongements syllabiques ainsi que des modalités interrogative et exclamative lorsqu'elles ont été clairement perçues par le transcripateur.

9 Les formes graphiques des insertions pronominales sont au nombre de 16 : *en, la, le, l', les, leur, lui, me, m', te, t', se, s', nous, vous, y*, auxquelles il faut ajouter les deux formes de la négation : *ne, n'*.

TaLTaC2), d'extraire les locutions de ce type. La grammaire locale opérationnelle utilisée pour la recherche avancée est la suivante :

«en CATSEM(ant)» OR «en CATSEM(insert) CATSEM(ant)» OR «en CATSEM(insert) CATSEM(insert) CATSEM(ant)» OR «en CATSEM(insert) CATSEM(insert) CATSEM(insert) CATSEM(ant)»

Un tel modèle permet de saisir les faux négatifs¹⁰. Par la suite, il faut procéder à un nettoyage des données ainsi extraites pour éliminer les attestations non vraies (faux positifs), c'est-à-dire les suites syntaxiques qui ne sont pas des gérondifs (i.e. *en avant*, *en tant*, *en autant* etc.) ou des gérondifs grammaticalisés (i.e. *en admettant que* etc.). Finalement il faut récupérer les vraies attestations qui pourraient être mal catégorisées en raison d'une imperfection dans la segmentation du corpus (i.e. *en enachetant un*). Après cette démarche, les gérondifs simples et avec insertions sont « lexicalisés » afin de permettre l'extraction de leurs concordances.

3.3. Procédure d'extraction des occurrences du participe présent

Une fois effectuée la saisie du gérondif, nous pouvons aborder l'analyse du participe présent. Il s'agit d'une forme simple, donc, a priori, facilement extractible par le biais d'une requête simple (*ant). Elle pose cependant un problème considérable lié à l'ambiguïté grammaticale des formes. Au-delà des fausses attestations que constituent les formes syntaxiquement non ambiguës qui ne sont pas des verbes (i.e. *tant*, *quant*, *pourtant*, *cependant*, *intendant*, *instant* etc.) et, en particulier, le nom et l'adjectif verbal orthographiquement distincts du participe présent (i.e. *suffocant*, *fatigant*, *intrigant*, *fabricant* etc.)¹¹, il existe de très nombreuses formes ambiguës (i.e. : *maintenant* : Adv ou PP^{ant} ; *montant* : Nom ou PP^{ant} ; *important* : Adj ou PP^{ant} ; *pendant* : Prép ou PP^{ant}). Dans ce cas, une alternative s'impose à nous : i) éliminer à la source ces formes au risque de perdre les quelques attestations du participe présent dont l'incidence est statistiquement faible (la probabilité d'être un participe présent étant inférieure à 5%¹²) ; ii) conserver ces formes et procéder à leur élimination après une analyse linguistique des concordances. C'est la deuxième solution que nous avons choisie, car elle rend visibles des cas qui peuvent être linguistiquement significatifs même s'ils sont quantitativement négligeables¹³.

Afin de pouvoir disposer d'un protocole exploitable sur d'autres corpus, nous avons établi une liste des non-verbes en /-ant/ non ambiguës, puis nous avons extrait du corpus les formes en /-ant/

10 Il est possible de trouver jusqu'à 2 clitiques entre /en/ et /-ant/ (*en l'y invitant*) ce qui porte à trois le nombre de places syntaxiques disponibles si l'on ajoute aux pronoms les formes graphiques de la négation (i.e. *en ne le leur disant pas*). Même s'il est peu plausible que toutes les places syntaxiques insérées soient saturées dans l'oral spontané, nous maintenons une grammaire opérationnelle à trois insertions car elle permet de récupérer des attestations qui sans cela n'auraient pas été extraites.

11 Il n'est pas superflu de relever que la norme orthographique qui établit pour certains verbes une distinction graphique entre le participe présent et l'adjectif verbal (*convainquant/convaincant* ; *fatiguant/fatigant*, *résidant/résident*) est complexe et qu'elle n'est pas systématiquement respectée par les scripteurs, y compris dans l'écrit formel de presse. Il peut donc être utile de contrôler au cas par cas ces occurrences afin de ne pas risquer d'éliminer en amont de vraies attestations.

12 Dans les corpus d'oral spontané par exemple la totalité des occurrences de *maintenant*, *important*, *restaurant*, *pendant* sont des non-verbes.

13 La désambiguïté grammaticale par le biais des concordances a montré par exemple que dans CFPP sur 36 occurrences de *amusant* et *vivant*, 3 seulement sont des PP^{ant} contre 33 occurrences d'adjectif.

n'appartenant pas à cette liste¹⁴, les formes en /-ant/ de gérondifs étant rendues inaccessibles en raison de l'opération préalable de « lexicalisation » (cf. § 3.2). Enfin, nous avons éliminé les fausses attestations restantes parmi lesquelles figurent les formes désambiguïsées adverbiales, adjectivales et nominales (i.e. *maintenant, marrant, courant* etc.).

4. Analyse comparée de l'usage du gérondif et du participe présent selon la modalité discursive

4.1. Distribution différentielle des gérondifs et participes présents

Le Tableau 1 présente les premiers résultats de notre analyse sur la distribution différentielle des gérondifs et participes dans les corpus de l'oral et de l'écrit. N représente le nombre d'occurrences de chacun des modes. V(FG) indique le nombre de types de formes graphiques distinctes actualisées. V(LE) représente le nombre de lemmes verbaux distincts employés au participe présent et au gérondif.

	PFC		CFPP		Corpus d'oral réunis		LH2010	
	V(FG)	N	V(FG)	N	V(FG)	N	V(FG)	N
Tot corpus	18 806	529 543	15 923	483 152	27 169	1 012 695	148 989	6 873 329
	V(LE)	N	V(LE)	N	V(LE)	N	V(LE)	N
Gér	72	200	66	194	105	394	1 282	9 657
PPant	12	104	24	108	28	205	1 901 *	23 239 *

Tableau 1 : Distribution des Gér/PP^{ant} dans les corpus d'oral et d'écrit. (*) valeur surestimée, cf. note 16.

Les deux corpus d'oral spontané, comparables dans leur taille (N) et dans le volume de leur vocabulaire (V(FG)), présentent une tendance uniforme dans la distribution des deux modes verbaux (N/Gér : 200 PFC, 194 CFPP ; N/PP^{ant} : 104 PFC, 108 CFPP). Il est intéressant d'observer que la différence de protocole de recueil des données de l'oral – entretiens longs (CFPP) et entretiens courts (PFC) – ne semble pas avoir d'incidence significative sur les phénomènes observés¹⁵. En effet, dans les deux corpus les gérondifs sont deux fois plus fréquents que les participes présents.

Afin de rapporter ces premières observations au comportement des deux modes verbaux dans le français écrit¹⁶, il nous a paru utile de réunir les deux corpus en un unique Corpus d'oral quantitativement plus comparable avec le Corpus d'écrit, puisque dans un rapport

14 Pour le moment cette liste comprend plus d'une centaine de formes, parfois rares telles que : *inélegant, sacripant, surpuissant, tout-venant, angl.warrant* etc.

15 Comme nous l'avons déjà dit au paragraphe 2.1, le corpus CFPP est constitué d'enregistrements longs (de 60 à 105 minutes). Les deux corpus oraux ayant approximativement la même taille, il s'ensuit que le nombre de locuteurs du CFPP est inférieur à celui du PFC. Nous pouvons donc faire l'hypothèse que la variable idiolectale a peu d'incidence sur la distribution des deux modes verbaux dans l'oral spontané.

16 Pour le corpus d'écrit, nous avons éliminé les types ambigus dont moins de 10% des occurrences étaient des participes (i.e. *concernant, militant, montant*, etc). Pour le moment nous avons limité ce contrôle manuel aux formes présentant au moins 50 occurrences, par conséquent les valeurs de N/PP^{ant} et de V(LE)/PP^{ant} du tab.1 sont surestimées.

de 1 : 7 (1 million à 6,9 millions d'occurrences). Ainsi notre analyse peut être menée selon une double modalité différentielle : la modalité discursive (oral / écrit) et les modes verbaux (gérondif / participe présent). Un bref examen du tableau 2 permet de constater que la tendance distributionnelle des deux modes verbaux à l'écrit est inversée par rapport à l'oral : l'écrit présente environ le double d'occurrences de participes présents par rapport au gérondifs ; l'oral, environ la moitié. Pour ce qui est de la variabilité lexicale attestée par ces deux formes à l'oral et à l'écrit, nous développons l'analyse ci-après (cf. § 4.3).

4.2. *Gérondifs et participes présents « propres » vs grammaticalisés*

L'analyse des concordances montre l'existence de deux types gérondifs et de participes présents : des formes que nous qualifions propres et des formes à base gérondive ou participiale qui sont le résultat d'un processus de grammaticalisation. Les premières ne sont pas figées, il s'agit de faits de discours : le participe présent ou le gérondif résultent d'une construction syntaxique de la part du locuteur ; les secondes sont des usages figés, il s'agit de faits de langue, à savoir d'usages de ces modes verbaux actuellement attestés en langue française comme adverbe, locution conjonctive ou prépositionnelle. Pour le recensement des formes grammaticalisées du gérondif nous faisons référence, sans le discuter ici¹⁷, à l'inventaire proposé par Halmøy (2003a : 77-79) qui parle tour à tour de figement et de grammaticalisation pour évoquer le fait qu'une forme gérondive puisse ne pas sembler totalement libre et qui compte parmi ces formes *en attendant*, *en attendant que*, *en passant*, *en passant à/par*, *en allant à/vers/sur*¹⁸. Pour les formes grammaticalisées du participe présent, nous nous appuyons sur les attestations recensées dans le dictionnaire de langue (Petit Robert 2008) de participes qui assurent le rôle d'une autre partie du discours que le verbe (i.e. *concernant*, *suivant*, *donnant-donnant*, *(professionnellement) parlant*, *soi-disant*, *étant donné*, *étant donné que*, etc.). Ajoutons que même si *sachant que* et *pensant que* ne sont pas attestées comme conjonction par le Petit Robert 2008, nous considérons toutefois ces formes comme grammaticalisées. Le tableau 2 présente la proportion des formes propres et des grammaticalisations pour chacun des modes verbaux par rapport au total des formes gérondives et participiales de l'oral et de l'écrit.

17 Voir notre étude antérieure (Floquet *et al.*, 2012) qui met en évidence l'usage répandu à l'oral de la forme prépositionnelle *en parlant de* (au sens de *à propos de*) laquelle n'est pas répertoriée par Halmøy (2003a : 79) parmi les formes grammaticalisées.

18 Précisons sur ce point qu'il n'est pas toujours aisé de départager les cas, très probablement en raison même du fait que, en tant que formes non finies du verbe, le gérondif et le participe présent sont à la lisière entre le figé et le non figé. Les verbes de mouvement, classés par Halmøy (2003a) parmi les formes grammaticalisées, nous semblent tout particulièrement appeler un approfondissement théorique et la formulation de critères précis de grammaticalisation. Pour l'heure nous optons pour une solution maximaliste qui consiste à considérer comme grammaticalisées toutes les occurrences de ces types pour nos calculs.

	PFC		CFPP		C. d' oral réunis		LH2010	
N Gérondif	182	91,0%	170	87,6%	352	89,3%	9.038	93,6%
N Gramm. à base gérondive	18	9,0%	24	12,4%	42	10,7%	619	6,4%
Total des formes gérondivales	200	100,0%	194	100,0%	394	100,0%	9.657	100,0%
Densité Gér - Occ 10 ⁻⁴	3,4		3,5		3,5		13,1	
N participe présent	53	51,0%	69	63,9%	122	57,5%	23.101	99,4%
N Gramm. à base participiale	51	49,0%	39	36,1%	90	42,5%	138	0,6%
Total des formes participiales	104	100,0%	108	100,0%	212	100,0%	23.239	100,0%
Densité PPant - Occ 10 ⁻⁴	1,0		1,4		1,2		33,6	

Tableau 2 : Densité des deux modes verbaux et de leurs grammaticalisations (oral vs écrit)

Le tableau 2 confirme en premier lieu le parallélisme distributionnel des deux modes sur les deux corpus d'oral (§ 4.1) : statistiquement, tous les 10 000 mots il existe une probabilité de rencontrer 3,4 (PFC) et 3.5 (CFPP) gérondifs propres contre 1,0 et 1,4 participes présents. Proportionnellement, l'usage des ces deux modes est beaucoup plus fréquent à l'écrit qu'à l'oral : la probabilité de rencontrer une forme propre du gérondif y est en effet plus de trois fois supérieure (13,1 (écrit) vs 3,5 (oral) gérondifs tous les 10 000 mots) et la probabilité de rencontrer une forme propre du participe présent, près de 28 fois supérieure à l'écrit qu'à l'oral (33,6 (écrit) vs 1,2 (oral) participes présents tous les 10 000 mots). Indépendamment de cet usage différentiel selon la modalité discursive, la confrontation des modes verbaux montre, dans l'écrit de presse, une tendance inverse à celle de l'oral spontané : si les participes présents sont statistiquement 2,6 fois plus nombreux que les gérondifs à l'écrit (33,6 PP^{ant} vs 13,1 Gér pour 10 000 mots), cet usage s'inverse à l'oral où les gérondifs sont presque trois fois plus nombreux que les participes (3,5 Gér vs 1,2 PP^{ant} pour 10 000 mots). Non seulement les participes présents sont moins utilisés que les gérondifs à l'oral, mais 42,5% d'entre-eux sont grammaticalisés, c'est-à-dire vraisemblablement produits comme des unités lexicales figées, tandis que seulement 10,7% de gérondifs le sont.¹⁹ Une proportion très importante des participes présents de l'oral spontané jouent donc une autre fonction grammaticale que celle du verbe, et cet élément milite en faveur de l'hypothèse que le participe présent comme forme propre est une forme rare à l'oral.

4.3. Aspects lexicaux et morphosyntaxiques des gérondifs et des participes présents

Pour ce qui est de la variété lexicale actualisée par les deux modes verbaux, le tab. 1 montre que le gérondif s'actualise globalement à travers une plus grande variété de verbes que le participe présent quoique dans des proportions différentes : 6 fois plus dans PFC et 3 fois plus dans CFPP. À l'écrit, la diversité lexicale des deux modes verbaux se situe dans un ordre de grandeur comparable, à la faveur toutefois du participe présent (1 901 vs 1 282). À l'oral au contraire le gérondif (V(LE) : 105) manifeste une variété lexicale environ quatre fois supérieure à celle

¹⁹ Selon cet axe également la tendance s'inverse à l'écrit avec 6,4% de grammaticalisations de gérondif contre 0,6% de participes présents.

du participe présent (V(LE) : 28). Nous faisons l'hypothèse que l'examen des lemmes²⁰ les plus fréquents associés aux deux modes et aux deux modalités discursives (cf. tab. 3) peut faire émerger un nouvel aspect de leurs usages différentiels.

Corpus d'oral réunis				LH2010			
Gér		PPant		Gér		PPant	
en disant	56	étant	84	en faisant	369	ayant	853
en étant	27	ayant	19	en mettant	139	étant	814
en faisant	21	sachant	14	en prenant	137	faisant	489
en rentrant	14	parlant	7	en étant	105	visant	483

Tableau 3 : Lemmes associés aux *gér/pp^{ant}* les plus fréquents à l'oral et à l'écrit.

Les lemmes actualisés à l'oral comme à l'écrit sont tous des verbes de haute fréquence à l'exception de *viser*, lequel soulève la question théorique déjà évoquée (cf. note 18) des formes grammaticalisées.²¹ À l'oral certains lemmes sont associés de manière privilégiée à un mode plutôt que l'autre : *dire* et *faire* « préfèrent » le gérondif au participe présent, (ce qui n'est pas le cas pour *faire* à l'écrit). Deux verbes de très haute fréquence – *être* et *avoir* – se distinguent par une tendance commune quelle que soit la modalité discursive : ils recouvrent à eux seuls les formes de PP^{ant} les plus fréquentes de l'oral et de l'écrit, leur forme gérondivale étant moins fréquente. Mais s'agit-il toujours de participes présents à part entière? Une première analyse sommaire de la distribution des PP^{ant} *étant* et *ayant* fait apparaître des différences d'emplois de ces formes qui obligent à reconsidérer leur catégorisation. Chose remarquable, l'analyse des concordances montre qu'au-delà des grammaticalisations (à base *étant*/à base *ayant*)²², on peut aussi avoir affaire à des formes *ayant/étant* + participe passé du type :

<i>pas § au moment avec le XX § euh je</i>	étant	<i>enfin étant arrivé en France très jeune je pense que j' n' ai pas</i>
<i>n' ai aucune expérience de ça + mais § mon vélo</i>	étant	<i>immobilisé depuis maintenant + oh mon dieu cinq jours</i>
<i>jm' en réjouis § j' crois tous les jours surtout</i>	ayant	<i>travaillé en banlieue de longues années hein vraiment</i>

[source : CFPP]

dont la catégorie morpho-syntaxique est autre que celle du participe présent (cf. *étant Parisien*, *n'ayant rien à proposer*) puisque *être* et *avoir* y jouent le rôle d'auxiliaire actif ou passif et non de prédicat sémantique plein. Cet élément est capital pour l'analyse sémantico-aspectuelle de ces formes puisque, comme le rappelle à juste titre Arnavielle (2003 : 40) « Dans le cadre de

20 Rappelons que la notion de lemme associée au gérondif implique par exemple que *en disant* regroupe *en m'disant*, *en se disant*, *en s'disant*.

21 La forme *visant* à couvrir plus de la moitié des occ. de *visant* à l'écrit. Sa fréquence jointe au fait que cette forme n'est pas attestée par le dictionnaire de langue comme locution prépositionnelle rend légitime un approfondissement théorique sur la question de la frontière entre forme libre et forme grammaticalisée.

22 Nous comptons 21 *étant* grammaticalisés pour l'oral, 113 pour l'écrit. À côté des liste « officielles » déjà évoquées de grammaticalisations – *ceci étant (dit)*, etc. – apparaissent de nombreuses structures complexes – du type *l'objectif étant de*, *V+comme étant* – dont la fréquence justifie selon nous qu'elles fassent l'objet d'une étude spécifique indispensable et à laquelle nous nous emploierons ultérieurement.

l'opposition aspectuelle inaccompli-accompli, à la forme simple s'oppose une forme composée cumulant la valeur sécante de l'auxiliaire et celle, résultative, du participe passé : 'voyant' 'ayant vu'; 'venant' 'étant venu'.²³ L'approfondissement ultérieur de cet aspect fondamental par une analyse linguistique comparée des contextes d'usages et des différents emplois du participe présent est donc indispensable pour que la description des usages des deux modes verbaux soit complète.

5. Interprétation des résultats et hypothèses provisoirement conclusives

Encore que de manière provisoire, nous pouvons nous essayer dans quelques conclusions nécessitant de plus amples confirmations empiriques. À la différence de l'italien où l'usage du *gerundio* ne varie pas de manière significative du point de vue diamésique, c'est-à-dire que sa présence à l'oral et à l'écrit est pratiquement identique (Giordano et Voghera, 2002 : 293)²⁴, le français présente un tableau radicalement différent : d'une part gérondif et participe présent sont proportionnellement peu nombreux à l'oral par opposition à l'écrit ; d'autre part leur usage selon la modalité orale spontanée vs écrite est inversé : à l'écrit les participes présents prédominent largement sur les gérondifs, à l'oral au contraire, c'est le gérondif qui a la préséance. Ces données, qui restent naturellement à reproduire et confirmer sur un corpus plus étendu, renforceraient donc l'hypothèse d'une disponibilité plus importante du gérondif dans l'oral spontané, ce qui pourrait expliquer pourquoi celui-ci tend à y recouvrir des emplois du participe présent (Floquet *et al.*, 2012) :

E2: Et alors maintenant, la retraite?

R: Ah, ben la retraite, c'est bon. Ça a été très dur quand même, ça a été très dur hein, j'ai eu six mois qui ont été très durs. Mais bon maintenant ça va, mais je pense toujours hein, j'y pense toujours, puis bon ben maintenant. Donc il y a la pétanque qui me prend beaucoup hein. **En étant** président de la société, donc ça fait quand même des années. [source : PFC, 42arf1_LG]

La prédominance du gérondif à l'oral pourrait être interprétée, en partie, comme une conséquence du thème traité²⁵, mais aussi comme le résultat de principes d'organisation linguistique de niveau supérieur et non comme le résultat de choix ou de modalités discursifs. En effet considérons la spécificité du français quant à la manière de lexicaliser le mouvement²⁶, le gérondif fait alors

23 Fort de cette observation Arnavielle soulève la question de l'impropriété des dénominations de *participe passé composé* et *participe présent composé* – qui respectivement, masquent la valeur aspectuelle sécante de l'auxiliaire ou lui attribuent au contraire trop d'importance – et propose pour y remédier la dénomination de *participe composé*.

24 Giordano et Voghera (2002) ont montré que, à l'oral comme à l'écrit, le *gerundio* est un mode verbal uniformément minoritaire en italien par rapport aux autres formes verbales : à l'oral 1,5% de tokens de *gerundio* par rapport au 53% des indicatifs et au 15,2 % des infinitifs ; à l'écrit, 1,6 % de tokens de *gerundio* par rapport au 46,5% de l'indicatif et au 16,5% de l'infinitif. Précisons toutefois que l'étude référencée porte sur des corpus moins sectoriels que les nôtres et comprennent pour l'oral, outre les discussions guidées et libres, des conversations téléphoniques, des monologues ainsi que des programmes radio et télévisés ; pour l'écrit, mis à part les journaux, le corpus comprend des textes provenant des manuels scolaires de l'école primaire, des pièces de théâtre, des romans et des scénarios.

25 Floquet *et alii* (2012) étant essentiellement une étude de nature statistique et morpho-syntaxique, l'analyse des aspects sémantiques y est peu développée.

26 À cet égard, Talmy (1985) distingue les langues lexicalisant au niveau du verbe la manière de mouvement tandis que le déplacement est pris en charge par un satellite du verbe, la particule verbale (*He ran away*) et les

figure de construction obligée, au sens où il n'existe pas de solution alternative dans cette langue, pour exprimer à la fois le déplacement et la manière de se mouvoir. Nous sommes alors tentés de faire l'hypothèse que lorsqu'il apparaît dans une configuration de type : verbe de mouvement orienté²⁷ + gérondif exprimant la manière de déplacement) telles que :

A : J' ai dû , faire deux brasses , et **je suis ressorti en courant** par , pour aller me mettre à l' abri du vent derrière une dune [source : PFC, 42aad1_GG.txt]

JP : il y avait une muraille et en cou/ **il est parti en courant** il s' appelait euh [source : PFC, 11ajp1_GG.txt]

la nature « libre » du gérondif est à relativiser, précisément parce que, dans ce cas, sa présence est comme régie par une loi pour ainsi dire « systémique » du français. Cela offre une nouvelle occasion de soulever l'importance d'une identification théoriquement fondée de la nature exacte – que nous avons appelée propre– de ces formes en /-ant/.

Notre analyse a montré, par ailleurs, que le participe présent est non seulement plus rare, à l'oral, mais que 1) dans près d'un cas sur deux et par le jeu de grammaticalisation, il est le constituant d'une autre partie du discours que le verbe ; 2) qu'il est souvent l'actualisation des auxiliaires *être* et *avoir*. Pour expliquer ce dernier usage à l'oral nous sommes tentés d'avancer deux hypothèses. La première résulterait du fait que le « participe composé » est un moyen adapté pour exprimer des rapports temporels complexes entre procès et états dans un régime d'« empilement », propre à la parole en construction, qui s'accommode mal des rapports syntaxiques hiérarchisés de l'hypotaxe ; la seconde, serait liée à sa capacité à concilier l'expression d'une richesse aspectuelle particulière et la relative simplicité d'usage d'une forme non finie.

Comment peut-on expliquer une telle opposition dans les usages de l'oral spontané et de l'écrit? La réponse est sans doute à chercher dans l'idée d'une grammaire distincte qui régirait chacune des modalités. Nous inscrivons notre réponse provisoire, dans le cadre général esquissé par Barra Jover (2004), et Massot (2010), Zribi-Hertz (2011) où le français oral est actuellement considéré comme un système de plus en plus indépendant du français normé (attesté à l'écrit et véhiculé par l'école). Il s'agit d'un modèle explicatif centré sur l'hypothèse diglossique selon laquelle le français contemporain devrait être décrit comme présentant deux types de grammaires en intersection : la grammaire standard protocolaire (GS) et des grammaires dialectales informelles (GD). La grammaire protocolaire est l'expression d'une adéquation des discours à une « norme standard », elle est « socialement attendue » et étroitement liée avec l'écrit, tandis que les grammaires dialectales régissent « une activité dans toute situation où le locuteur ne cherche pas à se conformer à la norme » (Zribi-Hertz, 2011 : 237-238). Selon cette perspective, le gérondif et le participe présent seraient des formes dont l'usage se raréfie dans les grammaires non normées à telle enseigne qu'on pourrait imaginer sinon leur disparition à moyen terme, au moins leur limitation à quelques verbes fréquents, donnant lieu à des gérondifs et participes appelés, selon notre hypothèse, à une grammaticalisation probable.

langues romanes qui font l'inverse. Ainsi en est-il du français (*Il est sorti en courant*) où le verbe prend en charge l'expression de la directionnalité du mouvement et des éléments périphériques, tel le gérondif, l'expression de la manière de mouvement.

27 Voir sur ce point Escoubas-Benveniste (2005).

Concernant l'écrit, nous pensons que la prédominance du participe présent sur le gérondif peut être interprétée à la lumière des études diachroniques du français²⁸. Lecointe (1997) a bien décrit l'emploi massif de formes nominales en /ant/²⁹ dans le français préclassique constatant d'une part le retard par rapport à l'Italie, où le style en *ant* se développe à partir de Boccace, d'autre part les grands écarts d'un auteur à un autre. Il s'agit donc d'un phénomène textuel fortement sensible aux modèles étrangers, surtout latins et italiens³⁰. Par ailleurs, comme le précise Lecointe (1997 : 11) « le lien du style en *ant* avec l'écrit est à certains égards une évidence : une telle forme d'imbrication est difficilement pensable à l'oral, si ce n'est dans des formes d'élocution emphatiques calquées sur l'écrit. »

Or, ce qu'il nous intéresse ici de souligner est le lien étroit et, semble-t-il, ancien entre le registre écrit et le participe présent. Même si à l'époque classique on constate une progressive décréue du phénomène en raison de nouvelles exigences de clarté et de recherche d'une prose qui soit moins détachée de l'oral (le style en *ant* étant désormais considéré comme affecté) force est de relever que nos données contemporaines semblent indiquer que le participe présent demeure encore aujourd'hui une marque du style écrit, poli et bien normé.

Avant toute conclusion hâtive, il importe tout d'abord de nous assurer au préalable de la reproductibilité de nos résultats en étendant notre enquête à d'autres corpus d'oral spontané et à d'autres types d'oral puis d'explorer de manière systématique les pistes que cette seconde étude ont permis de préciser quant aux différentes « natures » propres ou grammaticalisées de ces modes verbaux, à leurs différents emplois et à la valeur prédicative et sémantico-aspectuelle que revêtent ces deux formes selon la modalité du discours.

Références

- Arnavielle, T. (2003). Le participe, les formes en *-ant* : positions et propositions. *Langages*, 149: 37-54.
- Arnavielle, T. (2010). Le gérondif : nouvelle définition d'un objet étrange. *Cahiers AFLS*, 16(1): 6-24 (<http://www.afls.net/cahiersrevue.php>).
- Barra-Jover, M. (2004). Interrogatives, négatives et évolution des traits formels du verbe en français parlé. *Langue française*, 141: 110-125.
- Bolasco, S. (2010). *TalTac2.10, Sviluppo, esperienze ed elementi essenziali di analisi automatica dei testi*. LED.
- Branca-Rosoff S., Fleury S., Lefevre F. and Pires F. (2000). Discours sur la ville. Corpus de Français Parlé Parisien des années 2000 (CFPP2000). (<http://cfpp2000.univ-paris3.fr/>)
- Combettes, B. (2003). L'évolution de la forme en *-ant* : aspects syntaxiques et textuels. *Langages*, 149: 6-24.

28 Sur les problèmes de morphologie diachronique liée au participe présent, à l'adjectif verbal et au gérondif, voir l'étude exhaustive de Zanola (1999).

29 Lecointe (1997) traite pour l'essentiel du participe présent, le gérondif étant moins présent à cette époque-là. Sur ce dernier aspect voir Halmøy (2003b).

30 Dans la traduction du latin, on observe que le participe présent et l'ablatif absolu ainsi que participe passé des verbes déponents correspondent à un participe présent en français. Pour ce qui est du rapport avec l'Italie, nous pouvons ajouter que : «le développement du style en *ant* revêt donc le caractère d'un phénomène relativement rapide, qu'il est difficile de ne pas mettre en rapport avec l'influence italienne, chez des auteurs tous plus ou moins étroitement en rapport avec l'Italie» (Lecointe, 1997, 13-14).

- Durand J., Laks B. and Lyche C. (2009). Le projet : une source de données primaires structurées. In Durand J., Laks B. and Lyche C., editors, *Phonologie, variation et accents du français*. Hermès, pp. 19-61.
- Escoubas-Benveniste M.-P. (2005). Les prédicats de mouvement orienté source-cible. Thèse de Doctorat en linguistique française, non publiée.
- Floquet, O. (2011). Notes sur la traduction du «gerundio». *Plaisance*, Sezione Lingua, 213-228.
- Floquet, O., Escoubas-Benveniste, M.-P. and Bolasco, S. (2012). Sur le gérondif dans le français parlé et écrit, IIIème Congrès Mondial de Linguistique Française, Lyon.
- Gettrup, H. (1977). Le gérondif, le participe présent et la notion de repère temporel. *Revue Romane*, 12 : 211-270.
- Giordano R. and Voghera M. (2002). Verb system and verb usage in spoken and written Italian. In Morin A. and Sébillot P., editors, *JADT 2002 : 6èmes Journées internationales d'Analyse statistique des Données Textuelles*. Irisa-Inria, pp. 289-300.
(<http://lexicometrica.univ-paris3.fr/jadt/jadt2002/tocJADT2002.htm>)
- Halmøy, O. (2003a). *Le gérondif en français*. Ophrys.
- Halmøy, O. (2003b). Quelques remarques sur les formes gérondives dans le Gargantua et le Tiers livre de Rabelais. *Le français préclassique*, 7: 13-28.
- Herlsund, M. (2000). Le participe présent comme co-verbe. *Langue française*, 127: 86-94.
- Lecoite, J. (1997). Le style en «ant» au XIVE en France: conscience syntaxique et options stylistiques. *L'information grammaticale*, 75: 10-14.
- Massot, B. (2010). Le patron diglossique de variation grammaticale en français. *Langue française*, 168: 87-106.
- Meillet A. (1912). L'évolution des Formes Grammaticales. *Scientia* 12, 384-400, réimprimé in Meillet, A., editor (1948) *Linguistique historique et linguistique générale*, 1, Champion, pp. 130-149.
- Rossi-Gensane, N. (2011). Oralité, syntaxe, discours. In Detey S., Durand J., Laks B., Lyche Ch., *Les variétés du français dans l'espace francophone : ressources pour l'enseignement*. Ophrys, pp. 83-106.
- Talmy, L. (1985). Lexicalization patterns : Semantic structure in lexical form. In Shopen T., editor, *Language Typology and Syntactic Description*. Cambridge University Press, vol. 3 : 57-149.
- Zanola, M.T. (1999). La diachronie des constructions participiales dans une perspective sémantico-textuelle, *L'analisi linguistica e letteraria*. 2: 333-384.
- Zribi-Hertz, A. (2011). Pour un modèle diglossique de description du français : quelques implications théoriques, didactiques et méthodologiques, *Journal of French Language Studies*, 21(2): 231-256.

Dictionnaire

Le Nouveau Petit Robert de la langue française 2008, sous la direction de Rey-Debove J. et Rey A., Paris, Dictionnaires Le Robert. 2007 nouvelle édition.