

Portraits stylistiques des présidents de la Ve République (1958-2019)

Cyril Labbé¹, Dominique Labbé², Jacques Savoy³

¹ Laboratoire d'Informatique de Grenoble (Université Grenoble Alpes) - cyril.labbe@imag.fr

² PACTE (Université Grenoble Alpes) - dominique.labbe@umrpacte.fr

³ Université de Neuchâtel- jacques.savoy@unine.ch

Abstract

Over the past sixty years, eight presidents successively headed the Fifth French Republic (de Gaulle, Pompidou, Giscard d'Estaing, Mitterrand, Chirac, Sarkozy, Holland, Macron). After presenting the corpus of their speeches - more than 8,751 texts and 18.5 million labelled words - the style of each of them will be characterized by the richness of their vocabulary (diversity and specialization), personal pronouns and the length of their sentences. This will be an opportunity to underline the singularity of each president and, above all, to recall the multiple interests of large bases of labelled texts.

Keywords: Stylometry – French president – Vocabulary richness – Pronouns – Sentence length.

Résumé

En soixante ans, huit présidents se sont succédé à la tête de la Ve République française (de Gaulle, Pompidou, Giscard d'Estaing, Mitterrand, Chirac, Sarkozy, Hollande, Macron). Après avoir présenté le corpus de leurs discours - soit plus de 8 751 textes et 18,5 millions de mots étiquetés - le style de chacun des présidents sera caractérisé à l'aide de la richesse du vocabulaire (diversité et spécialisation), des pronoms personnels et de la longueur des phrases. Ce sera l'occasion de souligner la singularité de chaque président et surtout de rappeler les multiples intérêts des grandes bases de textes étiquetés.

Mots clés : Stylométrie – Présidents français – Richesse du vocabulaire – Pronoms – Longueur des phrases.

1. Introduction

Chaque locuteur se caractérise par un certain vocabulaire et un style qui lui est propre. Grâce à l'informatique, on dispose d'études sur le vocabulaire d'un auteur ou d'un groupe d'auteurs. En revanche, la "stylométrie" – étude statistique du style – reste relativement en friche (Savoy 2020). A l'aide de quelques exemples tirés d'un vaste corpus de textes, nous présenterons quelques indices qui pourraient prendre place dans cette "science du style".

Depuis l'avènement de la Ve République française (1958), huit présidents se sont succédé, pour onze mandats, à la tête de l'Etat (annexe 1). Nous avons recueilli pour l'instant 8 749 textes, soit 18,5 millions de mots. Tous ces textes ont été corrigés et étiquetés selon les principes énoncés par Muller (1977) (présentation du corpus dans (Arnold et al. 2016)).

Ces interventions publiques se sont déroulées dans des cadres très divers. Dès lors, outre la question du genre (allocutions, conférences de presse, discours électoraux, etc.), certaines caractéristiques de ces textes peuvent refléter l'influence des circonstances dans lesquels ils

ont été émis. Pour faire apparaître les caractéristiques propres au style de chaque président, il faudrait pouvoir neutraliser ces facteurs exogènes en étudiant des interventions présentées dans un cadre à peu près semblable. Par exemple, les discours sur l'*Etat de l'Union* que prononcent les présidents américains chaque année depuis la fondation des Etats-Unis (Savoy 2015; 2017).

Depuis 1959, le discours de vœux aux Français est le seul exercice auquel tous les présidents se sont livrés (annexe 2). Ces textes ont déjà été étudiés par Finiss-Boursin (1992), Leblanc (2016) et leurs équivalents en italien par Pauli & Tuzzi (2009). Ces discours appartiennent tous à un genre unique (l'allocution radio-télévisée) et ont été prononcés dans les mêmes conditions. Autrement dit, l'influence du genre et des conditions d'énonciation se trouve neutralisée. Il ne reste que la personnalité de chaque président, ses choix stratégiques et les circonstances politiques du moment.

Les 60 allocutions comptent au total 70 881 mots, soit une longueur moyenne de 1 243 mots, soit (plus ou moins) une dizaine de minutes. Pompidou avec 596 mots a été de loin le plus sobre, puis vient Mitterrand avec moins de 1 000 mots. A l'opposé, Macron, l'actuel président est de loin le plus prolixe (deux fois la moyenne).

Ces allocutions vont permettre d'étalonner les indices stylistiques avec la certitude qu'ils renvoient aux personnalités, aux choix de communication - ou à la période historique - et non aux effets de "genre". Nous commencerons par présenter quelques indices stylistiques relevés sur ces textes avant de les généraliser à l'ensemble des interventions.

2. Etat de la question et principaux indices

Dans la suite, le terme "vocable" ("word type") désigne l'unité minimale du lexique d'une langue, association d'une "forme canonique" et d'une catégorie grammaticale (par exemple : "pouvoir, verbe"). La "forme" ("word form") désigne la graphie sous laquelle apparaît un certain nombre de fois, dans le texte, le vocable considéré et le "mot" ("word token") est sa concrétisation à un emplacement précis du texte (pour cette dernière acception on rencontre aussi le terme "occurrence"). Nous désignons par N la longueur d'un texte (nombre de mots) et par V son vocabulaire (nombre de vocables différents).

2.1. La richesse du vocabulaire

La richesse mesure l'étendue du vocabulaire mobilisé pour la production du texte, soit le rapport entre V et N connu sous le nom de "type-token ratio" (Herdan 1960; 1966), (Guiraud 1969). La relation entre ces deux grandeurs n'est pas linéaire ni géométrique et ne peut être exprimée par une formule simple (Baayen 2001; 2008).

Aujourd'hui, la solution habituellement adoptée consiste à calculer la moyenne du nombre de vocables différents observés dans des échantillons de même longueur (typiquement 1 000 mots), sans recouvrement entre ces échantillons (Cossette 1994 ; Covington et al. 2010), d'où le terme de "moyenne mobile de la diversité lexicale" (LD Lexical Diversity).

Ce calcul présente quelques difficultés. Un certain nombre de mots (à la fin du texte) ne seront pas pris en compte dans LD car la longueur totale du texte ne sera généralement pas un multiple de 1 000 mots. De plus, comme toute moyenne empirique, LD devrait se voir associer un écart-type pour rendre compte de la fluctuation de la variable au long du texte – ou variation naturelle du phénomène dans un état stationnaire - et de l'incertitude due à l'échantillonnage (fonction inverse de la racine carrée du nombre de tranches extraites dans le

texte). Cette incertitude est nulle si l'on mesure les N-1 échantillons différents que l'on peut extraire d'un texte long de N mots (prélèvement exhaustif). Le calcul de ce prélèvement exhaustif est présenté dans Hubert & Labbé (1988; 1994) et dans Labbé & Labbé (2014). Pour une application au vocabulaire gouvernemental espagnol, voir Alvarez et al. (2000). A cette diversité moyenne, on associe donc un écart-type qui mesure les fluctuations effectives de la diversité dans le ou les textes étudiés, et d'autre part, une mesure de la spécialisation du vocabulaire, c'est-à-dire la proportion des vocables qui sont affectés à un passage spécifique par opposition aux vocables généraux utilisés quel que soit le thème. Cette proportion peut être nulle, voire quelquefois négative, quand le seul vocabulaire général est utilisé dans tout le corpus. Naturellement, cette spécialisation – si elle n'est pas nulle – subit, elle aussi, des variations au long du texte mais, pour l'instant, cette variabilité ne peut être estimée.

2.2. Autres indices stylistiques

De nombreuses autres mesures du style ont été proposées, voir par exemple (Biber et al. 2009), (Biber 2006), (Reppen et al. 2002), (Biber et al. 1999) et pour le discours politique (Hart et al. 2013). Nous retiendrons ici les pronoms comme indices de la relation à l'auditoire et à ce qui est dit. Cette relation est connue en France sous le nom de l' "énonciation de la subjectivité dans le langage" (Benveniste 1956 ; 1958; 1970), (Kerbrat-Orecchioni 1981), (Maingueneau 1994), (Pennebaker 2011), (Arnold & Labbé 2015). La densité des pronoms est l'un des principaux indices de cette énonciation de la subjectivité et de la relation qu'établit l'orateur avec l'auditoire. Principalement, cette relation peut être interpellative (je-vous) ou inclusive (nous).

Enfin, la longueur moyenne de la phrase : SL (Sentence Length). Cet indice peut être utilisé pour l'attribution d'auteur (Manion et al. 2004) ou la mesure de la lisibilité (combiné avec la longueur des mots). La plupart du temps, seule la longueur moyenne (MSL) est calculée. Cependant, l'extrême étalement de la distribution des longueurs, et son profil non gaussien, font que la moyenne n'est pas forcément l'indice le plus représentatif de ces longueurs (Labbé & Labbé 2018). On complétera ici cette moyenne par la médiale, ou "seconde médiane" (MISL) définie par rapport à la moitié du texte qui doit être couvert par des phrases de longueur inférieure à cette valeur et inversement. Dans le cas d'une allocution, cela signifie que l'auditeur entend pendant la moitié du temps des phrases de longueur supérieure à cette médiale. A ce sujet, il y a un accord pour considérer que les phrases très longues (20 mots ou plus) sont souvent difficiles à comprendre, plus encore à l'oral qu'à l'écrit (Richaudeau 1981, (notamment p. 178-181) et 1988). Un orateur qui utilise des phrases longues (et complexes) aura donc moins de chance d'être compris par son auditoire que celui qui se contente de phrases brèves et simplement construites.

3. Le style des vœux présidentiels

Etant donné le cadre contraint des vœux télévisés (genre, circonstances), on s'attend à ce que les indices soient peu différents entre les présidents (l'idée d'un modèle générique se trouve chez Finiss-Boursin (1992) mais elle est nuancée par Leblanc (2016)). En fait, c'est plutôt l'inverse qui s'est produit : les indices stylistiques reflètent tous des choix – ou des caractéristiques personnelles – nettement différenciés.

3.1. Diversité et spécialisation

Le tableau 1 présente les résultats obtenus sur chacune des onze séries de vœux. La diversité et la spécialisation sont exprimées pour mille mots. Les présidents sont classés en fonction de

leur diversité de vocabulaire, de la plus faible (Giscard d'Estaing avec 306 vocables différents en moyenne pour 1000 mots) à la plus recherchée (de Gaulle avec 418 ‰, soit +37% par rapport à Giscard).

Tableau 1. Diversité de spécialisation du vocabulaire des présidents (classement en fonction croissante de la diversité)

Mandats	Diversité (‰)	100 = Giscard	Spécialisation (‰)
1974-1981 Giscard d'Estaing	306	100	182
2007-2012 Sarkozy	315	103	244
2017-2019 Macron	321	105	236
1995-2002 Chirac	355	116	84
2012-2017 Hollande	356	116	108
1969-1974 Pompidou	360	118	47
Moyenne	361	118	81
2002-2007 Chirac	376	123	- 46
1965-1969 De Gaulle	381	125	104
1988-1995 Mitterrand	388	127	57
1981-1988 Mitterrand	394	129	- 74
1959-1965 De Gaulle	418	137	- 105

Les différences sont donc importantes mais peut-on affirmer qu'elles sont statistiquement significatives ? L'écart type associé à LD permet de répondre à cette question (tableau 2).

Tableau 2. Mesure de la variabilité de la diversité du vocabulaire et délimitation d'un intervalle de confiance associée à cette variable.

Mandats	Diversité	Ecart-type (σ)	D-	D+
1974-1981 Giscard d'Estaing	306	8,4	290	322
2007-2012 Sarkozy	315	8,5	298	332
2017-2019 Macron	321	8,5	304	338
1995-2002 Chirac	355	9,2	337	373
2012-2017 Hollande	356	8,8	339	373
1969-1974 Pompidou	360	8,9	343	377
Moyenne	361	9,0	343	379
2002-2007 Chirac	376	8,8	359	393
1965-1969 De Gaulle	381	9,4	363	399
1988-1995 Mitterrand	388	9,8	369	407
1981-1988 Mitterrand	394	9,3	376	412
1959-1965 De Gaulle	418	9,6	399	437

Rappelons que LD résulte d'un tirage exhaustif (tous les échantillons possibles). Par conséquent, il n'y a plus d'incertitude due à l'échantillonnage et l'écart-type mesure seulement la variabilité de LD dans le corpus étudié. On définit autour de LD, un intervalle de confiance à 5 %, dont les bornes supérieure (D+) et inférieure (D-) sont : $D+ = LD + 1,96\sigma$ et $D- = LD - 1,96\sigma$

Les diversités du vocabulaire de Giscard (306 ‰), de Sarkozy (315 ‰) et de Macron (321) se trouvent à l'intersection des trois intervalles de confiance qui sont associées à ces indices [304

– 322]. Il est donc impossible de rejeter l'hypothèse selon laquelle ces trois présidents ont, en conscience ou non, fait le même choix stylistique d'une faible diversité du vocabulaire. C'est pourquoi ils sont rassemblés dans un même cadre en haut du tableau 2. Le même constat peut être fait à propos, d'une part, du premier mandat Chirac, de Hollande et de Pompidou (cadre central) et, d'autre part, des seconds mandats de Chirac et de Gaulle et des deux mandats de Mitterrand (cadre supérieur). Le premier de Gaulle (1959-1965) étant "hors normes".

On a donc trois groupes correspondant sans doute à trois choix stylistiques. Quels sont ces trois choix ? Pour ses vœux, Giscard a adopté une expression proche de l'oral soutenu. Il a été imité par Sarkozy et Macron. Ces trois présidents ont également choisi de spécialiser leurs allocutions (aux thèmes de l'actualité et en n'hésitant pas à répéter certains vocables adaptés à ces thèmes). Bref, ils ont émis un discours presque familier pour expliquer leur politique (dominante pédagogique).

De Gaulle et Mitterrand ont fait un choix opposé ; un vocabulaire étendu (écrit littéraire) mais des propos généraux, voire hyper-généraux. Trois présidents (de Gaulle 1, Mitterrand 1 et Chirac 2) se trouvent d'ailleurs dans la situation assez rare d'une spécialisation négative.

Quant à Pompidou, Hollande et Chirac (durant son premier mandat et le début du second), ils n'ont pas vraiment choisi, se situant à la moyenne, entre les deux styles opposés symbolisés par Giscard et de Gaulle.

Enfin, Mitterrand reconduit les mêmes choix stylistiques pendant ses 14 ans au pouvoir. Chirac semble en changer au profit d'une expression plus soutenue mais cela tient à ce que, dès avant son grave accident vasculaire (septembre 2005), ses collaborateurs prennent de plus en plus de poids dans la rédaction de ses allocutions. Il reste donc un tournant chez de Gaulle qui, à partir de décembre 1965, semble renoncer à une forme très littéraire et hyper-générale pour adopter un style un peu plus simple et (probablement plus proche des problèmes du pays comme le montre le vocabulaire caractéristique de cette période (Arnold 2019)). On peut donc penser que le ballottage, lors de la première élection présidentielle au suffrage universel (décembre 1965) puis la faible victoire aux législatives de 1967 et enfin les événements de 1968 ont contraint le Général à changer de ton, du moins lorsqu'il s'adressait aux Français.

L'étude des autres indices stylistiques viennent enrichir ces constats.

3.2. Personnalisation et relation à l'auditoire

L'énonciation de la subjectivité dans le langage peut être mesurée à l'aide de la densité des différents pronoms personnels (tableau 3). Rappel : le vocable "je" s'actualise dans les textes sous les formes "j'", "je", "m'" et "me", etc.

Pour chaque pronom, deux colonnes donnent sa densité chez le président étudié et son poids comparé au président qui l'utilise le moins (par exemple, pour le "je", le plus sobre est le Général durant son premier mandat). Ces colonnes donnent l'étendue de la distribution et montrent que les différences sont importantes. Par exemple, les vœux de Macron sont les plus personnalisés et cette personnalisation excède de 66% (les deux tiers) celle du deuxième Chirac et encore de 48% (moitié plus) celle du premier mandat du Général. Pour le "vous", Giscard en emploie 14 fois plus que de Gaulle (7^e colonne). Quant au "nous" sa densité double de Giscard à Macron.

La relative dépersonnalisation du second mandat Chirac peut s'expliquer par des raisons déjà évoquées. En revanche, de Gaulle et Mitterrand reconduisent sensiblement les mêmes choix d'un mandat à l'autre. De Gaulle et, dans une moindre mesure, Chirac ont choisi de s'effacer

devant le "nous". Cette première personne du pluriel signifie "nous les Français" car "vous qui m'écoutez" n'est presque pas employé (voir (Labbé 1997)). Par exemple, de Gaulle prononce 4,3 "nous" pour un "je" et le "vous" est quasiment absent. La relation inclusive domine. En quelque sorte le président semble prêter sa voix à la collectivité nationale.

Tableau 3. Les pronoms personnels. Classement des présidents en fonction croissante de la personnalisation de leurs vœux.

Mandats	Pronoms pers. (‰)	Chirac2 = 100	Je (‰)	De Gaulle =100	Vous (‰)	De Gaulle =100	Nous (‰)	Giscard = 100
2002-2007 Chirac	43,0	100	8,9	262	6,4	492	11,4	106
1959-1965 De Gaulle	48,2	112	3,4	100	1,3	100	15,0	139
1965-1969 De Gaulle	50,3	117	3,9	115	1,9	146	19,7	182
1995-2002 Chirac	60,5	141	11,2	329	6,3	485	18,4	170
Moyenne	60,5	141	13,3	391	8,1	626	15,2	140
2007-2012 Sarkozy	61,1	142	20,1	591	9,0	692	14,5	134
2012-2017 Hollande	61,7	143	18,0	529	8,1	623	12,8	119
1988-1995 Mitterrand	64,9	151	15,4	453	7,3	562	11,1	103
1969-1974 Pompidou	67,1	156	14,4	424	13,1	1008	18,4	170
1981-1988 Mitterrand	68,0	158	14,4	424	8,0	615	13,7	127
1974-1981 Giscard	69,3	161	16,9	497	18,5	1423	10,8	100
2017-2019 Macron	71,2	166	19,8	582	9,6	738	21,1	195

Pour les autres présidents, seuls Pompidou et Macron donnent au "nous" une place apparemment supérieure à celle du "je", mais il s'agit, au mieux, de "moi et vous qui m'écoutez" et non pas de "nous les Français" comme chez de Gaulle voire Chirac. A part Macron, chez les présidents qui ont choisi de personnaliser leurs vœux, la relation "je-vous" domine. Elle l'emporte presque totalement chez Giscard. Chez Macron, beaucoup de "nous" sont des pluriel de majesté, les autres signifient "moi et le gouvernement" ou "moi et les autorités", rarement "moi et vous qui m'écoutez".

Comme pour la richesse du vocabulaire, on aboutit à des choix très contrastés pour un exercice apparemment formel. Ces choix sont-ils confirmés par les longueurs de phrases ?

3.3. Longueurs des phrases

Rappelons que deux indices sont retenus pour mesurer les longueurs de phrases, à savoir la moyenne (MSL) et la médiale (MISL), cette dernière partageant en deux l'intervention (tableau 4).

Là encore les différences de style sont considérables. La longueur moyenne varie du simple au double en passant de Chirac (premier mandat) à de Gaulle (2^e mandat) ; la longueur médiale de 1 à 2,25 entre les mêmes. On remarque également une tendance à l'alourdissement chez le Général et la tendance inverse chez Mitterrand (dont la phrase s'allège entre le second et le premier mandat).

Enfin, si l'on admet que toute phrase longue peut engendrer des problèmes de compréhension, on voit que, en dehors de Chirac et Hollande, les vœux des autres présidents contiennent, au moins pour la moitié du texte, des phrases de plus de 20 mots qui comportent un risque d'incompréhension.

Tableau 4. Longueurs moyennes (MSL) et médiales (MISL) des vœux de chaque mandat présidentiel (classement par longueur moyenne croissante).

Mandats	MSL (mots)	Chirac 1=100	MISL (mots)	Chirac 1 = 100
1995-2002 Chirac	16,4	100	21	100
2002-2007 Chirac	17,6	107	23	108
2012-2017 Hollande	18,4	112	23	112
1969-1974 Pompidou	22,4	137	28	136
2007-2012 Sarkozy	22,5	137	29	138
1988-1995 Mitterrand	22,5	137	32	152
Moyenne	23,8	145	31	150
1974-1981 Giscard	25,0	152	31	150
2017-2019 Macron	26,3	160	36	172
1981-1988 Mitterrand	28,9	176	36	173
1959-1965 De Gaulle	29,3	179	39	186
1965-1969 De Gaulle	32,9	201	47	225

En conclusion, les indices stylistiques font apparaître des contrastes importants entre les vœux des présidents mais soulèvent évidemment une question : s'agit-il de caractéristiques conjoncturelles ou de choix fondamentaux ? Dans ce cas, il faudrait retrouver ces caractéristiques dans les autres interventions présidentielles.

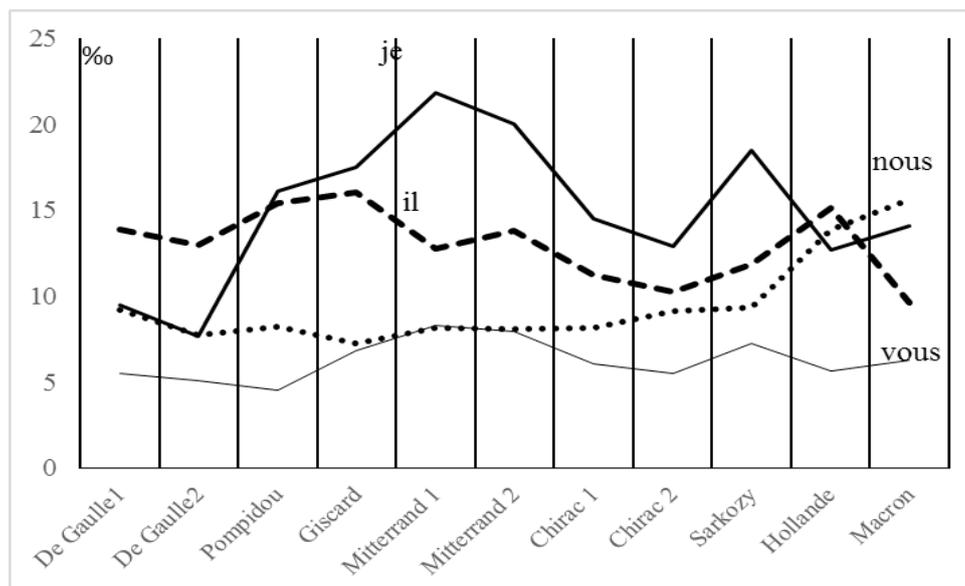
4. Application aux corpus entiers

La richesse du vocabulaire est dépendante des registres et, particulièrement, elle varie beaucoup entre l'oral et l'écrit. Cet aspect sera donc repris ultérieurement et nous limitons ici l'analyse à la personnalisation et aux longueurs de phrase.

4.1. La personnalisation

La figure 1 retrace l'évolution du poids des trois principaux pronoms personnels dans l'ensemble du corpus présidentiel (18,5 millions de mots) découpé en mandats.

Figure 1. Fréquence des pronoms personnels chez les présidents de la république depuis 1959 (pour mille mots).



Les densités et leurs évolutions sont semblables à celles qui ont été observées dans les discours de vœux.

Chez de Gaulle, l'évolution des quatre pronoms est corrélée ; le second mandat est moins personnalisé que le premier et tous les pronoms personnels subissent un recul à peu près semblable à partir de décembre 1965. Comme dans ses vœux, le "nous" du Général signifie "les Français". Après de Gaulle, "je" et "vous" évoluent dans le même sens et inversement à "nous" qui devient un synonyme de "moi et vous qui m'écoutez" et désigne, de plus en plus rarement, "nous les Français".

Dans les vœux, ce sont Sarkozy et Macron qui ont le plus employé "je". Pour l'ensemble des prestations présidentielles, Mitterrand est le plus "égocentrique", suivi par Giscard et Sarkozy puis Pompidou. A l'opposé, dans ses vœux comme dans l'ensemble de ses interventions, de Gaulle a été le plus sobre en pronoms de la première personne du singulier (du moins par rapport à ses successeurs).

4.2. Les longueurs de phrase

Tableau 5. Longueurs moyennes et médiales des phrases des présidents (ensemble de leurs interventions, classement par longueur croissante).

Mandats	Longueur moyenne (mots)	Sarkozy = 100	Longueur médiale (mots)	Sarkozy = 100
2007-2012 Sarkozy	22,5	100	28,3	100
1995-2002 Chirac	23,8	106	30	106
2002-2007 Chirac	23,9	106	29,8	105
1988-1995 Mitterrand	24,4	108	32,7	116
2012-2017 Hollande	24,2	108	30,8	109
1974-1981 Giscard	26,1	116	34,1	120
1981-1988 Mitterrand	26,5	118	35,6	126
Moyenne	27	120	35,1	124
1959-1965 De Gaulle	29,4	131	38,6	136
1969-1974 Pompidou	30	133	39	138
2017-2019 Macron	32,8	146	42,4	150

1965-1969 De Gaulle	33,6	149	44,6	158
---------------------	------	-----	------	-----

Par rapport au tableau 4, la hiérarchie est à peu près semblable ; avec des faibles longueurs, on retrouve Sarkozy, Hollande et Chirac. A l’opposé, de Gaulle et Macron. Entre les deux, Giscard et Mitterrand. Seul Pompidou ne retrouve pas sa place. D’une part, ce président a prononcé des vœux extrêmement brefs et d’autre part, ce corpus demeure le plus incomplet.

A part Pompidou, on peut en conclure que les styles des discours de vœux ne sont pas génériques ou conjoncturels mais appartiennent à des choix de communication propres à chaque président. On remarque en particulier, le choix pour plus de sobriété fait par Mitterrand dans son second mandat qui se retrouve à la fois dans ses vœux et dans l’ensemble de ses interventions ou l’évolution inverse chez le Général avec un net alourdissement de la phrase à partir de 1965 (sur ce phénomène, voir (Cotteret et Moreau, 1969), (Arnold, 2019)).

5. Conclusions

Pour un tableau complet, il faudrait mentionner d’autres aspects du style, notamment le poids des catégories grammaticales et, en particulier, la propension plus ou moins grande à mobiliser les verbes, les temps etc. Il faudrait également étudier les structures de phrases et bien d’autres aspects encore à découvrir.

Enfin, ces résultats devraient être confrontés à ceux obtenus par la statistique textuelle appliquée à ces discours présidentiels (notamment : Mayaffre 2004 et 2012 ; Mots 2016).

Ces réserves admises, notre étude débouche sur trois conclusions.

D’une part, en ce qui concerne les vœux présidentiels, les choix contrastés peuvent tenir aux circonstances, à des stratégies politiques et aux styles personnels mais, en tous cas, ils obligent à rejeter l’idée d’un genre ou d’un moule unique – le "discours de vœux" - qui s’imposerait aux locuteurs.

D’autre part, durant ces 60 ans, sept des huit présidents ont effectué des choix stylistiques qui se retrouvent dans l’ensemble de leurs interventions concernant le rapport à l’auditoire et aux thèmes développés – mesurés notamment par le jeu des pronoms personnels -, la longueur et la complexité des phrases

Troisième conclusion, le style peut faire l’objet d’une analyse statistique au même titre que le vocabulaire, à condition toutefois que l’on dispose de vastes corpus dépouillés avec soin et dont chaque mot aura été étiqueté sans erreur.

Remerciements

Les corpus des interventions des présidents français (1958-2018) sont disponibles sur demande auprès des auteurs. La plupart des interventions des présidents français depuis 1980 sont sur le site <http://www.vie-publique.fr/>. Les logiciels sont disponibles sur demande auprès de Cyril et Dominique Labbé.

Denis Monière (Université de Montréal) et Edward Arnold (Trinity College – Dublin) ont participé à la mise au point des outils de stylométrie utilisés dans cette communication.

Toutes nos recherches ont été réalisées sans aide publique ni mécénat.

Références

- Alvarez R. & Becue M. (2000). Le vocabulaire gouvernemental espagnol (1979-1996). *Mots*, 62-1, p. 32-47.
- Arnold E. (2019). Le vocabulaire et le style du Général de Gaulle. *Document Numérique*, 22, 11-30.
- Arnold E. & Labbé D. (2015). Vote for me. Don't vote for the other one. *Journal of World Languages*, 2(1), 32-49.
- Arnold E., Labbé C., & Monière D. (2016). *Parler pour gouverner : Trois études sur le discours présidentiel français*. Grenoble: Laboratoire d'Informatique de Grenoble (<https://hal.archives-ouvertes.fr/hal-01287657/document>).
- Baayen H.R. (2001). *Word Frequency Distributions*. Kluwer Academic Press, Dordrecht.
- Baayen H.R. (2008). *Analyzing Linguistic Data. A Practical Introduction Using R*. Cambridge : Cambridge University Press.
- Benveniste E. (1956). La nature des pronoms. In *Problèmes de linguistique générale*. Paris : Gallimard, 1966, p.251-257.
- Benveniste E. (1958). De la subjectivité dans le langage. In *Problèmes de linguistique générale*. Paris : Gallimard, 1966, p.258-265.
- Benveniste E. (1970), L'appareil formel de l'énonciation, *Langages*, 17, p 12-18.
- Biber D., Conrad S. & Reppen R.. (1999). *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber D. & Conrad S. (2009). *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Cossette A (1994). *La Richesse lexicale et sa mesure*. Paris-Genève: Slatkine-Champion.
- Cotteret J.-M. & Moreau R. (1969). *Le Vocabulaire du général de Gaulle*. Paris : Presses de la Fondation Nationale des Sciences Politiques.
- Covington M.A. & McFall J.D. (2010). Cutting the Goridan knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94-100.
- Finniss-Boursin F. (1992). *Les discours de vœux des présidents de la République, La France au fond des yeux*. Paris : Librairie générale de droit et de jurisprudence.
- Guiraud Pierre (1969). *Essais de stylistique*. Paris: Klincksieck.
- Hart R.P., Childers J.P. & Lind C.J. (2013). *Political Tone. How Leaders Talk and Why*. Chicago: The University of Chicago Press.
- Herdan G. (1960). *Type-Token Mathematics*. La Haye: Mouton.
- Herdan G, (1966). *The Advanced Theory of Langage as Choice and Chance*, Berlin-New York, Springer-Verlag.
- Hubert P. & Labbé D. (1988). A model of Vocabulary Partition. *Literary and Linguistic Computing*. Vol. 3, n° 4, p. 223-225.
- Hubert P. & Labbé D. (1994). Vocabulary Richness. *Communication au congrès de l'ALLC-ACH*, Paris: La Sorbonne. Reproduced in *Lexicometrica*, 0, 1997 (<http://www.cavi.univ-paris3.fr/lexicometrica/>).
- Kerbrat-Orecchioni C. (1981). *L'énonciation de la subjectivité dans le langage*. Paris, A Colin.
- Labbé C. & Labbé D. (2014). Was Shakespeare's Vocabulary the Richest? In Née Emilie, Daube Jean-Michel, Valette Mathieu, Fleury Serge (dir.). *Proceedings of the 12th International Conference on Textual Data Statistical Analysis*. Paris, p 323-336.
- Labbé C. & Labbé D. (2018). Les phrases de Marcel Proust. In Iezzi Domenica F., Celardo Livia, Misuraca Michelangelo. *Proceedings of the 14th International Conference on Statistical Analysis of Textual Data*. Roma: UniversItalia, 2018, p. 400-410.
- Labbé D. (1997). Le "nous" du général de Gaulle. *Quaderni di studi linguistici*. 4/5, 1998, p 331-354.

- Leblanc J.-M. (2016). *Analyses lexicométriques des vœux présidentiels*. Londres : ISTE éditions.
- Mainguenea D. (1994). *L'énonciation en linguistique française*. Paris, Hachette.
- Mannion D. & Dixon P. (2004). Sentence-length and Authorship Attribution: the Case of Oliver Goldsmith. *Literary and Linguistic Computing*, 19-4, p. 497–508.
- Mayaffre D. (2004). *Paroles de président. Jacques Chirac (1995-2003) et le discours présidentiel sous la Vème République*. Paris : Honoré Champion.
- Mayaffre D. (2012). *Nicolas Sarkozy. Mesure et démesure du discours (2007-2012)*. Paris : Presses de Science Po.
- Mots (2016). *Discours présidentiels et de présidentielles*. 112, 2016/3.
- Muller C. (1977). *Principes et méthodes de statistique lexicale*. Paris: Hachette.
- Pauli F., & Tuzzi A. (2009). The end of year addresses of the presidents of the Italian republic (1948–2006): Discourse similarities and differences. *Glottometrics*, 18, 40–51.
- Pennebaker J.W. (2011). *The Secret Life of Pronouns. What our Words Say About us*. New York: Bloomsbury Press.
- Popescu I.-I., Altmann, G. Grzybek P., Jayaram B. D., Köhler R., Krupa V., Mačutek J., Pustet R., Uhlířová L. & Vidya M.N. (2009). *Word Frequency Studies*. Berlin: De Gruyter Mouton.
- Reppen R., Fitzmaurice S. & Biber D. (2002). *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins Publishing Company.
- Richaudeau F. (1981), *Linguistique pragmatique*, Paris, Retz.
- Richaudeau F. (1988), *Ce que révèlent leurs phrases*, Paris, Retz
- Savoy J. (2015). Text clustering: An application with the *State of the Union* addresses. *Journal of the American Society for Information Science & Technology*, 66(8), 1645-1654.
- Savoy J. (2017). Analysis of the style and the rhetoric of the American presidents over two centuries. *Glottometrics*, 38(1), 55-76.
- Savoy J. (2020). *Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling*, Cham, Springer.

Annexe 1. Les interventions des présidents de la République française (1958-2019)

Mandats	Nombre de textes	Longueurs (mots)	Vocabulaire
1959-1965 De Gaulle*	304	281 096	7 494
1965-1969 De Gaulle*	155	129 396	5 930
1969-1974 Pompidou*	137	259 918	8 076
1974-1981 Giscard d'Estaing*	191	660 560	9 535
1981-1988 Mitterrand	1 590	3 363 824	20 861
1988-1995 Mitterrand	957	2 212 420	18 635
1995-2002 Chirac	1 305	2 444 858	19 834
2002-2007 Chirac	1 173	1 636 839	16 160
2007-2012 Sarkozy	1 074	3 221 250	21 597
2012-2017 Hollande	1 544	3 182 939	20 403
2017-2019 Macron**	319	1 040 496	14 819
Total	8 749	18 433 596	45 575

* En cours de complément grâce aux archives présidentielles déposées aux archives nationales

** L'année 2019 est en cours de dépouillement.

Annexe 2. Les allocutions de vœux des présidents de la Ve république française (1959-2019)

Mandats	Nombre de Discours	Longueurs (mots)	Longueur moyenne	Vocabulaire
1959-1965 De Gaulle	5	6 719	1 344	1 316
1965-1969 De Gaulle	4	4 674	1 169	1 113
1969-1974 Pompidou	5	2 982	596	748
1974-1981 Giscard d'Estaing	7	7 947	1 135	1 186
1981-1988 Mitterrand	7	5 985	855	1 146
1988-1995 Mitterrand	7	6 950	993	1 427
1995-2002 Chirac	7	8 993	1 285	1 369
2002-2007 Chirac	5	5 631	1 126	1 062
2007-2012 Sarkozy	5	6 419	1 284	1 140
2012-2017 Hollande	5	7 274	1 455	1 332
2017-2019 Macron	3	7 307	2 436	1 217
Total	60	70 881	1 243	