



**HAL**  
open science

## La richesse du vocabulaire politique : de Gaulle et Mitterrand

Dominique Labbé

► **To cite this version:**

Dominique Labbé. La richesse du vocabulaire politique : de Gaulle et Mitterrand. Sylvie Mellet et Marcel Vuillaume. Mots chiffrés et déchiffrés. Mélanges offerts à Etienne Brunet., Champion, pp.173-186, 1998, 978-2-85203-894-3. halshs-02217258

**HAL Id: halshs-02217258**

**<https://shs.hal.science/halshs-02217258>**

Submitted on 31 Jul 2019

**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.

Dominique Labbé.

Institut d'études politiques de Grenoble  
BP 45 38.402 Saint Martin d'Hères Cédex

## **"La richesse du vocabulaire politique : de Gaulle et Mitterrand".**

In Sylvie Mellet et Marcel Vuillaume. *Mots chiffrés et déchiffrés. Mélanges offerts à Etienne Brunet.*

Paris : Champion, 1998, p. 173-186.

Résumé : Mesure de la "richesse du vocabulaire" chez E Brunet. Application aux allocutions radio-télévisées de C. de Gaulle et F. Mitterrand. On propose de scinder la notion en deux dimensions : la diversité du vocabulaire et sa spécialisation. Les mesures confirment les valeurs obtenues avec l'indice de Brunet tout en les affinant. On peut alors isoler l'oral et l'écrit et opposer la préparation soignée et l'unité thématique chez de Gaulle au style oral et l'adaptation à l'événement chez Mitterrand.

Manuscrit de l'auteur.

Toute citation doit être faite à partir de l'ouvrage dirigé par S. Mellet et M. Vuillaume.

## LA RICHESSE DU VOCABULAIRE POLITIQUE DE GAULLE ET MITTERRAND

De Gaulle et Mitterrand sont considérés comme des hommes de culture dont le lexique aurait été fort étendu. Le premier doit sans doute cette réputation aux mots rares dont il parsemait ses interventions et le second à sa fréquentation des oeuvres littéraires et à son souci du mot juste qui se manifestait, lors de ses interventions publiques, par de fréquentes digressions "sémantiques". Des intuitions comparables, concernant le style et la culture des auteurs, se rencontrent fréquemment dans la critique littéraire. Depuis longtemps, la statistique lexicale cherche à les vérifier par des approches plus rigoureuses. Cela revient à se demander comment mesurer l'étendue du vocabulaire utilisé dans un texte ou dans une oeuvre. E. Brunet a proposé une solution originale à cette question avec son indice W qu'il a appliqué aux nombreux auteurs dont il a effectué le dépouillement<sup>1</sup>. Nous l'avons utilisé à notre tour sur les discours de C. de Gaulle et de F. Mitterrand durant leurs mandats présidentiels<sup>2</sup>. Nous présentons ci-dessous les résultats de ces mesures ce qui nous conduira à envisager des solutions nouvelles, largement inspirées des travaux pionniers de C. Muller et d'E. Brunet.

---

<sup>1</sup> L'indice W est présenté par Etienne Brunet dans *Le vocabulaire de Jean Giraudoux. Structure et évolution*, Genève-Paris, Slatkine-Champion, 1978. La formule est  $W = N^{V-a}$  avec N, nombre total d'occurrences ; V : nombre de mots différents. Au cours de son oeuvre, E. Brunet a proposé plusieurs valeurs pour "a". Nous avons retenu 0,185 qu'il utilise depuis son livre sur le vocabulaire de V. Hugo. Rappelons que la richesse du vocabulaire est d'autant plus forte que W est plus faible.

<sup>2</sup> Interventions radiotélévisées et conférences de presse du général de Gaulle entre 1958 et 1969 (201 900 occurrences) et de F. Mitterrand entre 1981 et 1988 (305 124 occurrences). Ces textes ont été lemmatisés selon la norme préconisée par C. Muller. Les données concernant F. Mitterrand ont été présentées dans notre ouvrage, *Le vocabulaire de F. Mitterrand*, Paris, Presses de la FNSP, 1990.

## *La richesse du vocabulaire*

E. Brunet a appliqué son indice W à un grand nombre d'œuvres — de Chateaubriand à Zola en passant par Giraudoux, Gracq, Hugo, Proust... — ainsi qu'au vocabulaire français de 1789 à nos jours. Ces expériences apportent une moisson de mesures et des conclusions intéressantes qui permettent de formuler plus précisément la problématique de la “richesse du vocabulaire”.

En premier lieu, E. Brunet a constaté une augmentation constante de la “richesse” lexicale avec le temps, augmentation qui tiendrait à un phénomène “démographique” propre à toute langue vivante : l'excédent des “naissances de mots” sur les “décès” engendre une “inflation lexicale”. Logiquement, on en tire que, plus un auteur est contemporain, plus ses textes ont de chances de contenir un vocabulaire étendu (puisque le stock de mots à sa disposition est supérieur à celui des générations précédentes). Autrement dit, il serait peu légitime de comparer des auteurs séparés par un trop grand nombre d'années, tout au moins lorsque le plus contemporain l'emporte sur le plus ancien. Dans ce cas en effet, on risque d'attribuer ce classement à la culture alors qu'il peut tenir à la différence d'étendue du lexique disponible aux deux époques. Autrement dit, la comparaison des richesses lexicales ne peut s'opérer que sur des auteurs plus ou moins contemporains. Dans le cas contraire, la conclusion n'est assurée que lorsqu'elle classe l'“ancien” devant le “moderne”. Par exemple, il en est ainsi de Hugo et de Proust : dans son ouvrage sur le vocabulaire de Zola, E. Brunet compare *Les travailleurs de la mer* et *Le temps retrouvé* et montre que le vocabulaire du premier est plus étendu que celui du second, ce qui va à l'encontre de certains clichés véhiculés à propos de Proust. On remarquera que telle est bien la conclusion à laquelle permet d'aboutir l'indice W appliqué à l'ensemble des corpus de Gaulle et Mitterrand : Le général fait jeu égal ou l'emporte légèrement sur son successeur (tableau B<sup>3</sup>).

Les travaux de E. Brunet permettent également de constater des différences assez nettes en fonction du genre auquel appartient l'œuvre analysée. L'étendue du vocabulaire se révélerait toujours supérieure dans la poésie et, plus particulièrement, dans la poésie en prose dont la richesse supplante nettement celle des œuvres en vers. De même, la prose littéraire

---

<sup>3</sup> Les tableaux sont placés à la fin de l'article.

l'emporte toujours sur les écrits techniques ou la correspondance. De ce point de vue, le dépouillement réalisé sur l'oeuvre de V. Hugo est particulièrement instructif puisque cet auteur prolifique a utilisé un grand nombre de genres au cours de sa longue vie littéraire.

On peut en déduire que les fluctuations dans la richesse du vocabulaire ne tiennent pas d'abord à la culture des auteurs mais à leur époque, aux genres des oeuvres ainsi qu'à certains choix stylistiques. Ces constatations amènent à reformuler la question de la richesse du vocabulaire : avant de mesurer celle-ci, il faut se demander si les oeuvres comparées sont vraiment comparables, c'est-à-dire si elles appartiennent à la même époque et au même genre littéraire.

Dans le discours politique, il est également possible de distinguer des "genres" (tableau I). L'étendue du lexique semble plus grande pour les messages de voeux et, plus largement dans les allocutions radiotélévisées, c'est-à-dire des messages rédigés et lus (ou appris par coeur). En revanche, les conférences de presse et les entretiens télévisés paraissent présenter une diversité lexicale plus faible. En fait, le principal discriminant semblerait être le partage entre écrit (allocutions) et oral (conférences, entretiens) bien qu'on puisse légitimement se demander quelle est la part des phrases préparées et apprises par coeur dans les propos tenus apparemment de manière spontanée lors de ces conférences et entretiens. Cette remarque complique la comparaison entre C. de Gaulle et F. Mitterrand. En effet, le premier a privilégié les allocutions et les conférences de presse, alors que le second a préféré multiplier les entretiens radiotélévisés : à partir de 1985, les conférences de presse élyséennes sont tombées en désuétude et les allocutions ont été limitées aux voeux de nouvel an. Telle est la raison pour laquelle, dans le tableau I, nous avons décomposé les deux corpus en sous-ensembles plus homogènes.

Au total, la meilleure performance est réalisée par de Gaulle lors de sa conférence de presse du 25 mars 1959 ( $W=9.9$ ). A ce propos, on notera que l'ensemble des indices obtenus sur les conférences de presse du général (tableau II) tendrait à prouver que ces textes appartiennent plus à l'écrit qu'à l'oral. A l'opposé, l'étendue du vocabulaire la plus faible se trouve aussi chez de Gaulle et, plus précisément, dans son entretien télévisé du 14 décembre 1965 ( $W=12.5$ ). En revanche, les performances de F. Mitterrand paraissent moins contrastées mais font également apparaître quelques différences entre l'"écrit" et l'"oral".

On remarquera enfin que l'étendue du vocabulaire des textes de C. De Gaulle et de F. Mitterrand — mesurée à l'aide de l'indice  $W$  — sont en général nettement plus faibles que celle observées par E. Brunet sur les oeuvres des grands écrivains du XIXe et du XXe siècle (sauf Zola). Mais les travaux de E. Brunet suggèrent qu'il ne faut pas imputer ce déficit aux différences de culture chez les auteurs ; il pourrait plutôt s'agir d'une caractéristique de la communication politique qui serait un genre mineur (du point de vue des qualités littéraires).

Au fond, la conclusion essentielle que l'on peut tirer des valeurs de  $W$ , c'est que C. de Gaulle et F. Mitterrand se valent approximativement du point de vue de l'étendue de leur vocabulaire et que cette étendue ne varie guère en fonction de la nature de l'exercice.

Toutefois, on peut se demander si l'indice  $W$  est suffisamment sensible au phénomène qu'il mesure ou, en d'autres termes, s'il n'est pas un peu trop stable. Pour vérifier cette intuition, on peut se tourner vers la méthode proposée par C. Muller<sup>4</sup>. Lorsque les textes à comparer sont de longueurs différentes, C. Muller préconise un "raccourcissement" du texte le plus long à la dimension du plus bref. On peut alors observer le nombre de mots différents contenus dans chacun d'entre eux et en déduire le plus "riche".

Si l'on baptise  $N$  la taille du plus grand et  $N'$  celle du plus petit, la formule proposée par Muller permet d'estimer le *nombre moyen de vocables différents contenus dans des échantillons de  $N'$  mots prélevés dans une urne contenant  $N$  mots* (avec  $N \gg N'$ ). Le prélèvement est effectué avec remise, c'est-à-dire que le contenu de l'urne (les  $N$  mots employés dans le corpus total) n'est pas modifié par les tirages successifs (telle est la raison pour laquelle il est préférable que  $N$  soit grand par rapport à  $N'$ ). Autrement dit, la formule de Muller est une excellente approximation de la loi hypergéométrique (même si elle a été improprement qualifiée par son auteur de "binomiale"<sup>5</sup>).

---

<sup>4</sup> Charles Muller, *Principes et méthodes de statistique lexicale*, Paris, Hachette, 1977, chapitre 20.

<sup>5</sup> La démonstration est donnée dans Pierre Hubert, Dominique Labbé, "Note sur l'approximation de la loi hypergéométrique par la formule de Muller" in Dominique Labbé, Philippe Thoiron, Daniel Serant, *Etudes sur la richesse et la structure lexicales*, Genève-Paris, Slatkine-Champion, 1988.

Cette définition suggère une solution réconciliant la perspective d'E. Brunet et celle de C. Muller : il suffit de donner la même taille à tous les échantillons prélevés sur chacun des textes comparés pour obtenir un indice unique. Pour permettre une saisie immédiate du résultat, nous avons proposé de donner systématiquement à N' la valeur de 1.000.

Cependant, l'utilisation de la formule de Muller se heurte à une difficulté qui a été souvent signalée : les valeurs calculées sont pratiquement toujours supérieures aux données observées. Ce phénomène peut atteindre des proportions importantes sur les grands corpus comme l'attestent notamment les travaux de Brunet. Evidemment, ce biais systématique rejaillit sur la comparaison des richesses du vocabulaire de plusieurs textes. Les plus longs sont "avantagés" et le sont d'autant plus que l'écart entre les données théoriques et les observations est important.

Ch. Muller expliqué ce phénomène par la *spécialisation lexicale* des différentes parties du texte en fonction du thème traité. Selon une expression d'E. Brunet, la spécialisation du vocabulaire forme des "grumeaux dans la pâte lexicale" et éloigne le texte du modèle théorique de l'urne. Il faut donc intégrer cette dimension supplémentaire dans la mesure de la richesse lexicale.

### *La spécialisation du vocabulaire*

L'écart entre les valeurs calculées grâce à la formule de Muller et les données observées permet de mesurer la spécialisation du vocabulaire : plus l'écart est grand, plus cette spécialisation est forte. L'indice de "partition" mesure cet écart ou, plutôt, il indique le poids relatif des deux sous-ensembles composant le vocabulaire d'un texte : d'une part, les mots spécialisés (qui sont employés uniquement pour traiter un thème donné) et, d'autre part, les mots "généraux" que l'auteur utilise quel que soit le thème abordé<sup>6</sup>.

L'indice est nul lorsqu'aucune spécialisation lexicale se fait sentir dans le texte étudié. Ce cas particulier correspond à la formule de Muller. A l'inverse, l'indice serait égal à 1 si tous les mots étaient spécialisés. Dans ce

---

<sup>6</sup> Pour la méthode de calcul : Pierre Hubert, Dominique Labbé, "Un modèle de partition du vocabulaire" in Dominique Labbé, Philippe Thoiron, Daniel Serant, *op cit.*

cas, V serait une fonction linéaire de N. En pratique, cette seconde situation est impossible. En effet, une fraction importante du vocabulaire est nécessairement “général” : mots outils, auxiliaires, verbes usuels, substantifs et adjectifs les plus courants... Telle est la raison pour laquelle on ne rencontre pas habituellement d'indices supérieurs à 0.4 dans un corpus relativement homogène. Autrement dit, lorsqu'un texte a été produit par un seul auteur à une date donnée, ou pendant une période relativement courte, il est peu probable que l'on rencontre dans le vocabulaire de ce texte, plus de 40% de mots spécialisés et donc moins de 60% appartenant au vocabulaire général.

Lorsque la valeur de l'indice est égale ou supérieure à 0.25, on peut considérer que la spécialisation lexicale est forte. Par exemple, l'indice atteint 0.39 dans les propos de F. Mitterrand lors du débat télévisé, entre les deux tours de l'élection présidentielle de 1981, et 0.35 chez son adversaire (V. Giscard d'Estaing). Or, les indices sont habituellement beaucoup plus faibles chez ces deux hommes ; ce qui permet de vérifier que la spécialisation n'est pas une caractéristique propre au style d'un auteur mais d'abord le produit de la situation d'énonciation. Lorsque la pression des autres participants à l'échange est forte, le locuteur se trouve contraint de spécialiser ses propos et de répondre précisément aux questions ou aux interpellations.

On remarquera que, tant chez de Gaulle que chez Mitterrand, la spécialisation la plus forte est atteinte dans des prestations orales et dans des situations d'échange. Pour C. de Gaulle, il s'agit d'abord de ses trois entretiens télévisés avec M. Droit en décembre 1965. Son ballottage au premier tour de l'élection présidentielle avait été une surprise et, après un temps d'hésitation, le président avait décidé d'utiliser son temps de parole en accordant trois entretiens qu'il n'avait pas préparés, contrairement à son habitude. Il avait le sentiment que son action était mal comprise — notamment à propos de la politique agricole commune — et s'il s'explique simplement, voire familièrement. En revanche, les deux derniers entretiens (de 1968 et 1969) ne présentent pas les mêmes caractéristiques, ce qui explique que la fusion des cinq émissions, en un seul corpus, engendre une sorte de "monstre" dont près des deux tiers du vocabulaire sont spécialisés. En fait, comme nous allons le voir à propos de la diversité du vocabulaire, ces textes appartiennent à des planètes différentes. Les trois premiers sont spontanés ; les deux derniers sont rédigés, appris par coeur et récités avec



tout le naturel dont était capable le général, mais, du coup, leur vocabulaire est faiblement spécialisé.

Dans le corpus Mitterrand on trouve également quelques entretiens dont les indices sont exceptionnellement importants. A chaque fois, la pugnacité de ses interlocuteurs et les circonstances particulières ont contraint F. Mitterrand à quitter les prudentes généralités pour s'expliquer avec plus de précision. L'indice maximal (0.35) est atteint lors de son entretien avec P. Alexandre sur RTL en novembre 1987, entretien qui a porté principalement sur des ventes clandestines d'armes à l'Iran qui auraient financé le parti socialiste. Au cours de cet entretien, le journaliste a relancé le F. Mitterrand sans concession. Le président s'est trouvé dans une situation comparable lors de son passage au magazine économique *l'Enjeu*, en septembre 1983, pendant lequel, devant des journalistes sceptiques, il a tenté de justifier la politique de rigueur, lancée en juin 1982, tout en prétendant poursuivre les réformes engagées en 1981.

Entre 0.10 et 0.25, la spécialisation est moyenne. C'est le cas de l'ensemble des deux corpus dont les indices sont proches de 0.1. On notera également que ces indices moyens se retrouvent quasiment à l'identique dans les conférences de presse et dans les entretiens. En fait, la tendance aux généralités est une caractéristique que l'on rencontre chez tous les hommes politiques, lorsqu'ils peuvent s'en tenir aux propos qu'ils ont préparés pour l'occasion.

En dessous de 0.10, la spécialisation du vocabulaire est faible. Cela peut signaler un choix conscient : celui de traiter le sujet sur un mode général. C'est le cas de toutes les allocutions de C. de Gaulle comme de F. Mitterrand. Un indice de spécialisation faible, sur des prestations "orales" signale donc que le texte a été rédigé et que le locuteur parvient à s'y tenir. C'est le cas de pratiquement toutes les conférences de presse du général de Gaulle ! Mais l'exemple le plus éclairant est fourni par les discours, prononcés les 31 décembre, pour présenter les vœux du président de la république à la nation. Chez C. de Gaulle comme chez F. Mitterrand, on observe une spécialisation légèrement négative, ce qui est un phénomène assez rare. Cette "hyper-généralité" est particulièrement révélatrice de la tendance à se réfugier dans les propos convenus qui est le propre de tous les grands leaders politiques contemporains. Mais comment faire autrement quand il faut s'adresser à tous sans fâcher personne ?

### *La diversité du vocabulaire*

Le modèle de partition permet donc de supprimer l'écart existant entre le profil théorique calculé avec la formule de Muller et le texte réel. Ainsi pouvons-nous calculer un indice vraiment indépendant de la longueur des textes comparés. Cet indice, noté  $V'(1000)$  dans les tableaux, mesure le nombre de vocables différents observés dans *toutes les tranches de 1.000 mots contigus* que l'on peut extraire d'un texte comportant au total  $N$  mots.

On verra dans les tableaux que les valeurs de cet indice sont cohérentes avec celles de  $W$ . Plus  $W$  est faible, plus on a de chance de rencontrer un vocabulaire étendu dans les tranches de mots prélevées aléatoirement sur le texte. Cependant, les écarts mis en évidence par l'indice de diversité sont plus importants et donc mieux lisibles. Ils permettent de trancher nettement en faveur du corpus de Gaulle dont la diversité du vocabulaire dépasserait de 11% celle du corpus Mitterrand : en moyenne 363 vocables pour 1.000 mots chez le premier contre 328 chez le second (tableau I).

L'indice de "diversité du vocabulaire" mesure la propension de l'auteur à diversifier son expression, à fuir la répétition ou, à l'inverse à réutiliser les mêmes mots. Naturellement, le résultat sera très influencé par les conditions d'émission du message. Lorsqu'on parle sans préparation, on choisit entre deux inconvénients : le risque d'une répétition ou une pause nécessaire à la recherche du mot juste, mais qui risque d'indisposer l'auditoire. En revanche, à l'écrit, la relecture permet d'éliminer les répétitions par le recours au synonyme ou à la périphrase. C'est pourquoi, l'oral se signale également par un excédent des verbes les plus usuels, notamment le verbe "faire", et de quelques mots "passe-partout" comme "chose".

Les indices les plus forts — supérieurs à 400 vocables pour 1000 mots — sont atteints par le général de Gaulle, par exemple lors de ses allocutions de vœux ou dans celles qu'il a prononcées lors des grandes batailles politiques de ses onze ans de pouvoir et notamment pendant la guerre d'Algérie<sup>7</sup>. On remarquera que cette performance est également atteinte dans quatre conférences de presse (en particulier celle de novembre 1959 où il présente l'autodétermination en Algérie et celle du 9 septembre 1965 au cours de

---

<sup>7</sup> Jean-Marie Cotteret et René Moreau ont proposé de baptiser ces allocutions de combat des "allocutions-appels" (*Le vocabulaire du général de Gaulle*, Paris, Presses de la FNSP, 1969).

laquelle il dresse le bilan de son premier septennat et annonce implicitement sa candidature à l'élection de décembre). Ceci prouve que, non seulement le général apprenait par coeur ses réponses aux questions — qu'il connaissait à l'avance — mais aussi qu'il soignait particulièrement ces textes. En effet, F. Mitterrand, au cours de son premier septennat, n'atteint qu'une seule fois l'indice 400 (allocution de voeux du 31 décembre 1982).

Les propos spontanés tenus sans réflexion préalable dépassent rarement l'indice 300. Pourtant des valeurs aussi faibles se rencontrent rarement dans les corpus politiques "oraux". Chez F. Mitterrand, l'entretien avec P. Alexandre et le magazine *l'Enjeu*, que nous avons déjà mentionnés, présentent les valeurs les plus basses avec respectivement 280 et 290 vocables pour 1.000 mots. Chez C. de Gaulle, seuls les trois entretiens de décembre 1965 ont une diversité inférieure à 300. Le troisième, avec 262 mots est, de loin, le texte le plus "pauvre" qu'il nous ait été donné de dépouiller dans le discours politique contemporain car, non seulement, il s'agit d'une authentique improvisation mais aussi d'un choix stylistique conscient.

Au fond, l'émetteur d'un message a le choix entre quatre "registres" possibles qui ne sont pas exclusifs : la science, la pédagogie, la polémique et la poésie. La communication scientifique fuit la diversité, le vocabulaire doit y être univoque et les mots sont répétés autant que de besoin (mais il n'est pas nécessaire de revenir sur ce qui est déjà démontré). En revanche, le pédagogue répète pour attirer l'attention de l'auditoire et souligner les points clefs de l'exposé. Comme le scientifique, il fuit la diversité du vocabulaire parce qu'elle peut "brouiller le message". Les propos du pédagogue sont donc peu divers, lexicalement parlant. A l'inverse, aux yeux du poète, un vocabulaire dépouillé est synonyme de monotonie et de pauvreté du message. Le polémiste se trouve un peu dans une situation semblable : le discours de combat est un discours à forte diversité lexicale même s'il ne dédaigne pas quelques répétitions d'emphase.

Le général de Gaulle spécialisait assez bien ses interventions. Certaines étaient à tonalité pédagogique — notamment les conférences de presse dont, sauf quelques exceptions, l'indice est assez bas — d'autres avaient une visée nettement polémique. Par exemple, dans les allocutions de combat, dont nous parlions plus haut, l'indice moyen atteint 410, ce qui est considérable si l'on songe que nombre de ces textes ont été rédigés en quelques heures dans des situations de grande tension.

### *Le problème des grands corpus*

Les indices de spécialisation et de diversité sont réellement significatifs lorsqu'ils sont mesurés sur des ensembles assez homogènes. En revanche, les résultats obtenus sur de vastes corpus hétérogènes doivent être accueillis avec prudence. En effet, effectuer ce genre de mesures sur un vaste corpus revient à postuler implicitement que les textes rassemblés n'en forment plus qu'un, ce qui n'est pas toujours légitime. On en trouvera une confirmation dans un phénomène quasi-systématique. Les indices — que ce soit le W de Brunet ou les P et V' présentés ici — ne sont jamais “transitifs” : les valeurs obtenues sur l'ensemble s'écartent toujours nettement de la moyenne des valeurs obtenues sur chaque partie de cet ensemble. Un exemple proprement caricatural est fourni par les conférences de presse. Chez C. de Gaulle, la diversité de l'ensemble — mesurée avec V' comme avec W — est inférieure à la diversité de toutes les conférences prises individuellement sauf une (tableau II) ; la situation est exactement inverse chez F. Mitterrand (tableau III).

Parfois, la réunion de plusieurs textes, appartenant apparemment au même genre, engendre un véritable monstre. Le cas le plus typique est fourni par les cinq entretiens télévisés du général de Gaulle (tableau IV). Apparemment, il s'agit du même type d'émission avec le même journaliste. Pourtant, les indices de diversité et de spécialisation signalent que tout change entre 1965 et 1968-69 : spontanéité et improvisation d'un côté, mise en scène et texte appris de l'autre.

L'explication de ces perturbations réside d'abord dans ce qu'on pourrait appeler la "dimension cachée" qui indique, pour chaque mot de la langue, l'espace minimal devant séparer ses occurrences. Cet espace peut être de quelques unités — pour les mots les plus courants (articles, conjonctions, prépositions, pronoms, verbes auxiliaires...) — ou s'étendre au texte entier pour des vocables rares dont la répétition passerait pour une faute de goût. Naturellement, cette dimension ne jouera plus, ou presque plus, quand on change d'oeuvre ou quand il s'écoule plusieurs mois entre deux interventions orales. Comme il se passait au moins six mois entre chaque conférence du général de Gaulle, la fusion de ces textes constitue un être plus pauvre que chacun d'entre eux pris individuellement.

Mais un second phénomène joue en sens inverse : les changements historiques introduisent de nouveaux thèmes et des mots nouveaux tout en déclassant les anciens. C'est ce qui s'est produit lors des conférences de presse de F. Mitterrand (tableau III). La première portait essentiellement sur les réformes entreprises après mai 1981 ; la seconde sur la politique internationale après le fastueux sommet de Versailles ; la troisième sur le "plan de rigueur" et le retour à l'orthodoxie économique et financière ; enfin la dernière — tenue à la veille des élections législatives de mars 1986 qui verront la première "cohabitation" entre un président de gauche et une majorité de droite — constitue une sorte de bilan politique de la législature et annonce le repli du président sur sa fonction. A chaque fois, un vocabulaire nouveau s'impose, le précédent passant au second plan ou disparaissant.

Autrement dit, les changements de situation dictaient ceux du vocabulaire de F. Mitterrand alors que chez le général de Gaulle, l'unité thématique transcendait les événements tout au moins quand il se livrait à un travail d'élaboration important. Probablement en aurait-il été de même pour F. Mitterrand si celui-ci était parvenu à s'imposer une discipline suffisante. En effet, la baisse continue de la diversité du vocabulaire de 1981 à 1985 (tableau III) révèle une préparation de plus en plus médiocre. Sans doute est-ce la raison pour laquelle, après 1985, F. Mitterrand a renoncé à ces conférences de presse.

En définitive, il est sans doute préférable de limiter l'usage de l'indice W — comme de V'(1000) et de P — à des corpus homogènes, tant du point de vue du genre que de la chronologie. D'autres mesures sont à imaginer pour mesurer l'effet du temps et des changements thématiques sur le contenu du vocabulaire.

En conclusion, nous espérons avoir montré que la problématique de la "richesse lexicale" ne mérite pas l'oubli dans lequel elle est progressivement tombée. En effet, elle semble actuellement condamnée par le développement de l'analyse des "données textuelles". En effet, cette dernière travaille sur le texte "brut", sur les formes graphiques sans aucune intervention préalable contrairement à la lemmatisation préconisée par C. Muller et utilisée dans la plupart des travaux d'E. Brunet. Etant donné la très

grande variabilité des graphies, l'incohérence des frontières entre les mots, et faute d'une norme commune de dépouillement, la comparaison devient difficile entre différents textes et, *a fortiori*, entre plusieurs auteurs traités par divers chercheurs... Les travaux d'E. Brunet montrent pourtant que d'autres voies sont possibles et qu'elles peuvent apporter beaucoup à la connaissance des oeuvres littéraires ou politiques.

Dominique Labbé  
Institut d'études politiques de Grenoble

Tableau I. Données d'ensemble concernant la richesse du vocabulaire dans les discours de C. de Gaulle (1958-1969) et F. Mitterrand (1981-88)

		N	V	W	V'(1000)	Spécialisation
Allocutions :	C. de Gaulle	71 547	4 135	11,0	391	0,052
	F. Mitterrand	8 996	1532	10,4	394	-0,028
dont Voeux	C. de Gaulle	12 883	1 921	10,3	423	- 0,080
	F. Mitterrand	5 977	1 147	10,6	396	- 0,084
Conférences :	C. de Gaulle	115 037	5 325	10,8	348	0,156
	F. Mitterrand	51 237	3 708	10,7	346	0,110
Entretiens	C. de Gaulle	18 416	2 018	11,1	212	0,630
	F. Mitterrand	188 911	6 268	11,1	335	0,122
Ensemble	C. de Gaulle	201 900	6 473	11,1	363	0,115
	F. Mitterrand	305 124	7 700	11,2	328	0,100

Tableau II. Diversité et spécialisation du vocabulaire dans les conférences de presse du général de Gaulle (1958-1969)

Date	N	V	W	V'(1000)	Spécialisation
23 octobre 1958	3 431	866	10,3	384	-0,031
25 mars 1959	1 820	603	9,9	378	0,277
10 novembre 1959	6 494	1 233	10,5	404	-0,160
5 septembre 1960	6 989	1 227	10,7	365	0,022
11 avril 1961	6 772	1 231	10,6	354	0,080
5 septembre 1961	5 447	1 129	10,4	345	0,197
15 mai 1962	6 100	1 111	10,8	351	0,077
14 janvier 1963	7 080	1 303	10,5	376	0,048
29 juillet 1963	6 877	1 383	10,2	402	0,027
31 janvier 1964	7 930	1 524	10,1	362	0,189
23 septembre 1964	6 035	1 284	10,1	387	0,102
4 février 1965	6 758	1 425	10,0	377	0,171
9 septembre 1965	8 234	1 576	10,1	402	0,023
21 février 1966	6 348	1 329	10,1	386	0,092
28 octobre 1966	7 635	1 473	10,2	345	0,245
16 mai 1967	6 727	1 391	10,1	409	0,030
27 novembre 1967	8 137	1 595	10,0	378	0,132
8 septembre 1968	6 223	1 307	10,1	405	0,000
Ensemble des conférences	115 037	5 325	10,8	348	0,156

Tableau III. Diversité et spécialisation du vocabulaire dans les conférences de presse de F. Mitterrand (1981-1985)

Date	N	V	W	V'(1000)	Spécialisation
24 septembre 1981	17 629	2 193	10,5	355	0,129
9 juin 1982	13 318	1 840	10,6	342	0,183
4 avril 1984	10 390	1 530	10,8	319	0,248
21 novembre 1985	9 900	1 540	10,7	312	0,288
Ensemble des conférences	51 237	3 708	10,7	356	0,110

Tableau IV. Diversité et spécialisation du vocabulaire dans les cinq entretiens télévisés du général de Gaulle (1965-1969)

Date	N	V	W	V'(1000)	Spécialisation
14 décembre 1965	3 272	545	12,5	281	0,124
15 décembre 1965	3 190	569	12,1	269	0,238
16 décembre 1965	2 934	548	12,0	262	0,322
7 juin 1968	7 709	1 017	10,5	361	0,094
10 avril 1969	4 311	986	10,4	390	-0,010
Ensemble des entretiens	18 416	2 018	11,1	212	0,630