



**HAL**  
open science

## La répartition des mots dans le vocabulaire présidentiel (1981-1988)

Pierre Hubert, Dominique Labbé

► **To cite this version:**

Pierre Hubert, Dominique Labbé. La répartition des mots dans le vocabulaire présidentiel (1981-1988).  
Mots: les langages du politique, 1990, 22, pp.80-88. halshs-01744995

**HAL Id: halshs-01744995**

**<https://shs.hal.science/halshs-01744995>**

Submitted on 27 Mar 2018

**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.

Pierre HUBERT

Ecole des Mines de Paris

Fontainebleau.

Dominique LABBE

CERAT-Institut d'études politiques

Grenoble

## **La répartition des mots dans le vocabulaire présidentiel (1981-1988)**

Article paru dans *Mots*, n° 22, mars 1990, p. 80-88.

### **Résumé**

Les mots employés par F. Mitterrand lors de son premier septennat sont analysés sous l'angle de leur répartition, c'est-à-dire de leur localisation dans les interventions du président. L'indice de répartition met en évidence deux vocabulaires caractéristiques. Le vocabulaire habituel, employé quelles que soient les circonstances, éclaire le rôle du président selon F. Mitterrand. Les mots qui appartiennent au vocabulaire circonstanciel sont localisés en certains points du corpus. Ils révèlent l'impact de certaines crises ou de préoccupations importantes mais passagères.

Mots clés : Mitterrand, répartition, localisation, vocabulaire circonstanciel, vocabulaire habituel.

### **Abstract**

Words Allocation in the Presidential Vocabulary (1981-1988). The words used by president Mitterrand during his first term in office are analyzed from the point of view of their "allocation", that is to say their localization in the corpus. The "allocation index" reveals two vocabularies: everyday words, used in any circumstances, reveal the Mitterrand's conception of the presidential function ; circumstantial vocabulary, composed of words used in some parts of the corpus, shows the impact on the discourse of situations of crisis or temporary, but important, preoccupations.

Key-words : Mitterrand, allocation of words, localization of words, circumstantial vocabulary, everyday vocabulary.

Manuscrit des auteurs. Toute citation doit se faire à partir de la revue.

Les mots les plus employés par les hommes politiques sont des mots banaux. L'index hiérarchique issu du dépouillement des interventions radio-télévisées du premier septennat de F. Mitterrand n'échappe pas à la règle<sup>1</sup>. Cette liste contient d'ailleurs un grand nombre de mots outils (articles, adverbes, pronoms...) sans véritable signification propre. Le sentiment de banalité vient de l'utilisation intensive qui en est faite par tous : il est impossible de les associer à l'un plutôt qu'à l'autre des acteurs politiques. A l'inverse, le terme rare ou inattendu restera en mémoire. Ainsi dans le vocabulaire du général de Gaulle<sup>2</sup>, les mots les plus fréquents ne l'évoquent pas particulièrement mais tout le monde se souvient du "quarteron de généraux en retraite", du "Volapük" ou du "machin" par lesquels le général a vilipendé les officiers putschistes d'Alger, l'Europe supranationale ou l'ONU... Apparemment aucune des formules de F. Mitterrand n'a connu un sort aussi favorable. Cependant, nous avons pu constater que beaucoup de personnes gardent en mémoire quelques scènes télévisuelles en particulier son dialogue avec Y. Mourousi à propos du "parler verlan"<sup>3</sup>. Ces formules ont été retenues parce qu'elles étaient totalement inattendues dans la bouche du général de Gaulle ou de F. Mitterrand. Certes la lexicométrie, parce qu'elle est quantitative, n'est pas *a priori* qualifiée pour traiter ce phénomène qualitatif ; nous voudrions cependant montrer qu'elle n'est pas totalement désarmée. Jusqu'à maintenant la lexicométrie s'est surtout attachée à la notion de fréquence admettant plus ou moins un postulat implicite selon lequel les mots possèdent, en n'importe quel point du texte, une probabilité d'emploi égale à leur fréquence relative dans l'ensemble. Ce postulat n'est pas vérifié : nous avons montré par ailleurs que le vocabulaire de F. Mitterrand change beaucoup dans chacune des quatre parties du septennat. Certains mots sont très utilisés dans une des périodes et complètement ou quasiment absents des trois autres. Comment chiffrer ce phénomène sur l'ensemble du corpus<sup>4</sup> et quelle signification lui prêter ?

---

<sup>1</sup> Voir ci-dessous le tableau 1 qui donne les cent vocables les plus utilisés dans le vocabulaire du septennat. La constitution du corpus et le dépouillement des textes sont présentés au début de notre article "Des réformes à la cohabitation. Les quatre périodes du premier septennat Mitterrand" dans ce même numéro.

<sup>2</sup> D'après MM. Cotteret et Moreau, il s'agit de : France, pays, république, Etat, monde, peuple, nation, progrès, paix, avenir (classés par ordre décroissant de fréquence. Jean Marie COTTERET, René MOREAU, *Le vocabulaire du général de Gaulle*, Paris, Presses de la FNSP, 1969, p 7.

<sup>3</sup> Première émission "ça nous intéresse monsieur le président" sur TF1 le 28 avril 1985.

<sup>4</sup> Les notions de "spécificités" et de "banalité" dépendent du découpage effectué ou choisi par l'observateur et ne permettent pas d'obtenir une mesure unique sur l'ensemble du corpus.

Nous proposons ci-dessous le calcul d'un indice qui répond à la première interrogation (comment mesurer ?). Dans l'analyse du vocabulaire présidentiel, ce calcul ajoute à la fréquence d'emploi, une dimension nouvelle : la répartition du mot<sup>5</sup>. Par cette mesure la lexicographie peut accéder à la structure même du discours.

L'indice de répartition varie entre zéro et un<sup>6</sup>. Une valeur proche de 1 signifie que le mot est employé pratiquement quelles que soient les circonstances ; on peut considérer qu'il appartient au vocabulaire habituel du locuteur. Plus l'indice est proche de zéro plus le mot est d'utilisation circonstancielle ; il appartient moins à l'auteur qu'à la situation d'énonciation ou au thème dont il parle : beaucoup de noms propres, de dates, de chiffres sont dans ce cas... Nous proposons de parler d'un vocabulaire circonstanciel ou localisé. Dans le corpus étudié, l'indice varie effectivement entre 0,8 pour le mot le plus régulier (*piéd*) et 0,1 pour le mot le plus localisé, c'est-à-dire le substantif *lycéen* (tableaux 2 et 3).

Nous présentons les résultats de l'expérience pour les mots les plus fréquents et pour les mots les plus habituels et les plus circonstanciels du vocabulaire de F. Mitterrand (tableaux 1, 2 et 3). Cette application montre que la notion présente un double intérêt. D'une part le calcul donne une mesure synthétique de la localisation des mots dans le vocabulaire du septennat fournissant un point de comparaison objectif entre les différents mots employés par le président sans avoir à prendre en compte des découpages plus ou moins arbitraires. D'autre part, l'indice permet d'identifier les régularités et les accidents et d'en tirer quelques conséquences sur l'importance de tel ou tel thème dans le discours présidentiel entre 1981 et 1988.

---

<sup>5</sup> Les linguistes ont baptisé répartition d'un mot dans un texte l'ensemble de ses localisations ou encore de ses "adresses" comme disent les informaticiens. Cf. notamment à ce sujet Charles MULLER, *Principes et méthodes de statistiques lexicales*, Paris, Hachette, 1977, p 55-66 et Pierre LAFON, *Dépouillements et statistiques en lexicométrie*, Paris-Genève, 1983, p 50-54.

<sup>6</sup> Les instruments classiques de la statistique sont peu utiles pour traiter de la plus ou moins grande dispersion d'un mot dans un texte. Si l'on raisonne sur les distances entre les occurrences, on constate une très forte dispersion de celle-ci autour de la moyenne (l'écart type dépasse souvent la moyenne) et une distribution fortement asymétrique. Dans de tels cas l'écart type, même normé, perd beaucoup de sa signification. D'où la nécessité de recourir à d'autres calculs.

Sans entrer ici dans le détail de la définition et du calcul de notre indice de répartition nous signalerons seulement qu'il est égal à la probabilité empirique d'apparition du mot considéré dans un segment quelconque du texte de longueur égale au quotient de la longueur totale du texte par la fréquence du mot considéré. Dans l'hypothèse d'une régularité parfaite, cette longueur sera la période selon laquelle le mot apparaîtra tout au long du texte. L'indice de répartition sera alors égal à un. A l'opposé, si toutes les occurrences du mot considéré sont contiguës, l'indice sera égal à zéro. Le calcul de l'indice est présenté en détail en annexe de : Dominique LABBE, Normes de saisie et de dépouillement des textes politiques. *Cahier du CERAT* n° 7. Grenoble : CERAT-IEP, avril 1990.

Tableau I Indice de répartition des cent mots les plus fréquents (ensemble du corpus Mitterrand, classement par fréquence décroissante)<sup>7</sup>.

Vocables	Fréquence	Indice	Vocables	Fréquence	Indice
<i>le (article)</i>	29 599	0,67	pays	742	0,47
<i>de</i>	20 964	0,67	comme	733	0,60
<i>être</i>	10 995	0,66	si (conjonction)	729	0,61
<i>je</i>	8 902	0,53	notre	710	0,49
avoir	7 713	0,63	celui	692	0,55
<i>à</i>	7 228	0,63	savoir (verbe)	672	0,59
<i>que (conjonction)</i>	5 547	0,63	vingt	666	0,47
un (article)	5 530	0,62	très	627	0,56
<i>et</i>	5 516	0,66	penser	610	0,55
ce (pronom)	5 260	0,60	là	609	0,60
il	4 976	0,59	grand (adjectif)	602	0,58
ne	4 665	0,62	même (adverbe)	575	0,61
qui	4 075	0,62	aussi	566	0,63
pas (adverbe)	4 042	0,61	leur (article)	563	0,51
ce (article)	2 723	0,63	deux	560	0,54
en (préposition)	2 683	0,61	tout (adverbe)	528	0,62
on	2 587	0,54	gouvernement	523	0,52
que (pronom)	2 326	0,61	non	523	0,56
pour	2 257	0,58	beaucoup	519	0,55
dans	2 187	0,62	donc	509	0,64
se	2 131	0,60	croire	500	0,57
le	2 076	0,60	autre (pronom)	491	0,60
nous	2 047	0,46	peu	485	0,56
<i>mais</i>	1 973	0,66	parler	472	0,56
dire	1 945	0,60	où	467	0,56
vous	1 907	0,53	monsieur	452	0,50
y	1 745	0,57	certain (adjectif)	446	0,56
faire	1 668	0,60	entendre	446	0,60
cela	1 633	0,59	quand	441	0,56
pouvoir (verbe)	1 491	0,60	autre (pronom)	439	0,62
par	1 297	0,56	<i>Français</i>	425	0,50
plus	1 288	0,58	parce que	425	0,58
falloir	1 286	0,55	alors	417	0,63
tout (déterminant)	1 238	0,61	politique (substantif)	405	0,47
<i>France</i>	1 219	0,54	chose	391	0,57
cent	1 200	0,44	premier	391	0,53
sur	1 176	0,57	<i>Europe</i>	390	0,34
bien (adverbe)	1 165	0,61	voir	388	0,57
ils	1 132	0,49	français (adjectif)	387	0,52
en (pronom)	1 043	0,60	sans	368	0,57
mille	957	0,44	déjà	366	0,60
vouloir	918	0,60	problème	365	0,52
mon	869	0,51	venir	358	0,59
son (article)	868	0,52	question	356	0,53
avec	842	0,60	prendre	349	0,57
neuf (numéral)	821	0,44	an	348	0,53
aller	786	0,56	temps	345	0,60
ou	777	0,59	tout (pronom)	345	0,54
devoir (verbe)	747	0,58	depuis	334	0,61
quatre	745	0,47	République	330	0,48

<sup>7</sup> Pour permettre de retrouver plus facilement les mots cités en exemple, ceux-ci ont été placés en italiques dans le texte et dans les tableaux.

## Le vocabulaire habituel

La stabilité d'un certain nombre de mots, qui est révélée par un indice de répartition élevé, peut recevoir trois interprétations complémentaires.

On pense d'abord à l'emploi obligatoire, en français comme dans toute langue, d'un certain nombre d'outils : articles, prépositions, adverbes, auxiliaires... Ces mots outils devraient donc retrouver en tête du vocabulaire habituel. Le tableau 2 impose une interprétation nuancée : l'article *le* ne survient qu'en 83<sup>ième</sup> position, la préposition *de* en 90<sup>ième</sup> et le verbe *être* en 99<sup>ième</sup>. Le tableau 1 confirme cette intuition : les mots-outils les plus fréquents ne présentent pas des indices de répartition remarquablement élevés et ils sont dépassés par ceux de nombreux mots pleins. Cela montre que l'emploi des déterminants, pronoms et autres prépositions n'est pas fixé *ne varietur* en langue mais peut varier de façon importante en fonction de la situation d'énonciation et du contenu du message. Cependant cette première intuition n'est pas totalement fautive : les fameuses formes fonctionnelles (*de, le, et, que, à, mais...*) ont des indices relativement élevés - habituellement supérieurs à 0,6 - ce qui atteste une relative stabilité...

La seconde explication possible est de nature exogène : le retour régulier de certaines circonstances amène celui de tel ou tel mot. L'examen du tableau 2 ne laisse guère de place à cette interprétation puisque la plupart des mots qui composent cette liste ne peuvent être rattachés à une circonstance particulière. Mais la chose n'est pas totalement invraisemblable comme le prouve, par exemple, le mot *table* que le président emploie 17 fois au cours de son premier septennat avec une grande régularité (l'indice est égal à 0,67)<sup>8</sup>. L'examen du contexte montre qu'il s'agit toujours de la *table de négociation*, l'utilisation de ce syntagme survenant à l'occasion de crises internationales qui ont elles-mêmes ponctué les sept ans : négociations sur les euromissiles, guerres du Liban, du Tchad ou entre l'Irak et l'Iran. Bref la stabilité de la formule tiendrait avant tout aux conflits internationaux récurrents. Cependant, on peut aussi considérer que l'emploi de ce syntagme manifeste d'abord la manière dont le président conçoit le règlement des conflits. On pourra alors relire le tableau 2 et trouver plusieurs autres mots qui confirmeraient cette intuition (notamment *réconciliation* ou *arbitrage entre les parties prenantes*).

Dès lors, il faut bien admettre que, pour le vocabulaire habituel, les causes de cette stabilité sont à rechercher principalement dans les habitudes de langage, la personnalité et l'état

---

<sup>8</sup> Cet indice signifie que le mot sera présent dans les deux tiers des segments de longueur N/F composant le corpus.

d'esprit du président. Un mot régulier appartient à la couche la plus profonde, au stock lexical qui sert pratiquement en n'importe quelle circonstance. Pour le premier septennat de F. Mitterrand, le tableau 2 nous indique de quoi se compose ce stock.

On trouve d'abord des expressions banales dont le mot le plus régulièrement employé durant le septennat (*pied*) fournit un exemple type. Dans le corpus, on trouve ce mot dans des formules comme : "partir d'un bon pied", "mettre sur pied", "être à pied d'oeuvre", "traîner les pieds", "avoir les pieds sur terre", etc. Ici l'habitude, les tics de langage fournissent la seule explication vraiment satisfaisante. Il en est de même pour les adverbes usuels : *malheureusement, récemment, spécialement, justement, essentiellement...*

Le vocabulaire habituel se compose ensuite de verbes comme *exposer, (se) comporter, résumer, préciser, excuser, rechercher, accorder, (se) tourner, traduire, confier, conclure, admettre...* Aucun de ces verbes n'a une fréquence très forte mais ils ont tous comme caractéristique d'être beaucoup employés à la première personne. Ils soulignent d'abord combien la parole est une clef de la fonction présidentielle selon F. Mitterrand : le président *expose, précise, accorde que, résume, confie, propose*, etc. Secondairement, il effectue des actions symboliques : il se *comporte, ose (déclarer que...), se tourne (vers... pour dire...)*, il *invite (à faire)*, etc. L'examen des substantifs les plus réguliers renforce ces intuitions puisque un grand nombre d'entre eux se rapportent à la conversation (*précision, considération, critique, leçon...*) ainsi qu'au rôle du président : *arbitrage, réconciliation, appel, message* ou encore *refus*. F. Mitterrand rappelle volontiers qu'il lui appartient de *dégager les (grandes) orientations, de rechercher le consentement (national)*<sup>9</sup> ou encore qu'il est le garant de l'indépendance nationale, de l'unité de la nation, de la constitution... Il est caractéristique que le seul nom propre qui apparaisse dans ce vocabulaire usuel soit Elysée. Il désigne à la fois le siège de la présidence et la fonction présidentielle. Il est, en quelque sorte, le noyau du vocabulaire habituel du président lors de son premier septennat.

---

<sup>9</sup> Le président refuse d'utiliser "consensus" (le mot était très employé par Giscard d'Estaing) mais il dit *rechercher le consentement (national)*. Là encore se dévoile l'importance que F. Mitterrand attache aux mots. De même, le président résiste à "rigueur" ou à "cohabitation" mais accepte *gestion rigoureuse* ou *coexistence...*

Tableau 2 Les cent mots les plus réguliers du corpus Mitterrand (classement par indice de répartition décroissant).

Rang	Vocable	Fréquence	Indice	Rang	Vocable	Fréquence	Indice
1	<i>pied</i>	11	0,79	51	être (subst.)	12	0,69
2	<i>exposer</i>	14	0,78	52	dégager	12	0,69
3	<i>scientifique (adj.)</i>	11	0,78	53	derrière	28	0,69
4	<i>camp</i>	14	0,77	54	rejoindre	12	0,69
5	<i>comporter</i>	17	0,76	55	comparer	10	0,69
6	<i>prenant</i>	13	0,75	56	conquérir	15	0,69
7	<i>arbitrage</i>	10	0,75	57	<i>message</i>	11	0,69
8	<i>précision</i>	16	0,75	58	<i>traduire</i>	11	0,69
9	<i>naissance</i>	10	0,75	59	<i>justement</i>	20	0,69
10	<i>considération</i>	10	0,74	60	fournir	24	0,68
11	distance	14	0,74	61	rude	15	0,68
12	<i>oser</i>	16	0,74	62	note	11	0,68
13	revanche	10	0,74	63	distinction	12	0,68
14	<i>malheureusement</i>	15	0,74	64	ambition	20	0,68
15	<i>résumer</i>	12	0,73	65	abri	10	0,68
16	vaincre	13	0,73	66	<i>efforcer</i>	30	0,68
17	<i>critique</i>	18	0,73	67	aggraver	10	0,68
18	<i>leçon</i>	10	0,73	68	<i>confier</i>	18	0,68
19	préoccupation	12	0,73	69	<i>conclure</i>	13	0,68
20	<i>récemment</i>	30	0,72	70	<i>discours</i>	29	0,68
21	<i>préciser</i>	17	0,72	71	interpréter	11	0,68
22	<i>excuser</i>	10	0,72	72	amical	10	0,68
23	ampleur	10	0,72	73	air	27	0,68
24	prévu	17	0,72	74	souvenir (subst.)	14	0,68
25	<i>réconciliation</i>	11	0,72	75	<i>admettre</i>	22	0,68
26	modeste	10	0,71	76	total (subst.)	23	0,68
27	<i>bref (adj.)</i>	14	0,71	77	garantir	13	0,67
28	<i>spécialement</i>	30	0,71	78	faciliter	15	0,67
29	<i>rechercher</i>	15	0,71	79	plupart	53	0,67
30	<i>accorder</i>	27	0,71	80	ancien	46	0,67
31	user	11	0,71	81	soi	34	0,67
32	sort	19	0,71	82	<i>table</i>	17	0,67
33	sien	10	0,70	83	<i>le (article)</i>	29 599	0,67
34	<i>tourner</i>	22	0,70	84	amener	11	0,67
35	véritable	19	0,70	85	relais	10	0,67
36	journée	10	0,70	86	inventer	19	0,67
37	priver	11	0,70	87	danger	20	0,67
38	<i>essentiellement</i>	20	0,70	88	habiter	13	0,67
39	dominer	24	0,70	89	<i>inviter</i>	37	0,67
40	source	10	0,70	90	<i>de</i>	20 964	0,67
41	extraordinaire	18	0,69	91	<i>proposer</i>	57	0,67
42	<i>Elysée</i>	23	0,69	92	tradition	22	0,67
43	<i>appel</i>	27	0,69	93	constituer	20	0,67
44	<i>consentement</i>	17	0,69	94	donnée	29	0,67
45	nature	29	0,69	95	espoir	21	0,67
46	concours	13	0,69	96	dès	104	0,67
47	<i>garant</i>	10	0,69	97	tomber	30	0,67
48	malgré	13	0,69	98	compliquer	12	0,67
49	refus	19	0,69	99	<i>être</i>	10 995	0,67
50	<i>orientation</i>	11	0,69	100	classique (adj.)	11	0,66



Tableau 3 Les cent mots les plus irréguliers du corpus Mitterrand (classement par indice de répartition décroissant).

Rang	Vocable	Fréquence	Indice	Rang	Vocable	Fréquence	Indice
1837	<i>Europe</i>	390	0,34	1887	<i>reconnaissance</i>	10	0,28
1838	<i>Strasbourg</i>	15	0,34	1888	<i>colonel</i>	14	0,28
1839	<i>Pologne</i>	22	0,34	1889	<i>Pisani</i>	11	0,28
1840	<i>communiste (adj.)</i>	17	0,34	1890	<i>Nouvelle-Calédonie</i>	33	0,28
1841	<i>Palestinien</i>	20	0,34	1891	<i>Varsovie</i>	11	0,28
1842	<i>Orsay</i>	18	0,34	1892	<i>palestinien (adj.)</i>	30	0,28
1843	<i>langue</i>	27	0,33	1893	<i>croissance</i>	64	0,28
1844	<i>Irak</i>	32	0,33	1894	<i>musée</i>	25	0,28
1845	<i>rassemblement</i>	15	0,33	1895	<i>Eurêka</i>	18	0,27
1846	<i>Union soviétique</i>	84	0,33	1896	<i>pershing</i>	34	0,27
1847	<i>Kadhafi</i>	15	0,33	1897	<i>radio</i>	41	0,27
1848	<i>entrepreneur</i>	15	0,33	1898	<i>immigré (subst.)</i>	28	0,27
1849	<i>Espagne</i>	49	0,33	1899	<i>Afrique</i>	106	0,26
1850	<i>taxe</i>	19	0,33	1900	<i>endettement</i>	27	0,26
1851	<i>producteur</i>	35	0,33	1901	<i>Renault</i>	16	0,26
1852	<i>candidat</i>	42	0,33	1902	<i>Canaques</i>	20	0,26
1853	<i>outré-mer</i>	31	0,32	1903	<i>Liban</i>	65	0,25
1854	<i>fusée</i>	36	0,32	1904	<i>sidérurgie</i>	32	0,25
1855	<i>métropole</i>	24	0,32	1905	<i>missile</i>	37	0,25
1856	<i>redresser</i>	13	0,32	1906	<i>Dreux</i>	12	0,25
1857	<i>superpuissance</i>	16	0,32	1907	<i>africain (adj.)</i>	47	0,25
1858	<i>épargne</i>	32	0,32	1908	<i>Lorraine</i>	35	0,25
1859	<i>prime</i>	11	0,32	1909	<i>Louvre</i>	26	0,24
1860	<i>agriculteur</i>	37	0,32	1910	<i>nationalisation</i>	57	0,24
1861	<i>arabe (adj.)</i>	50	0,31	1911	<i>Philippe</i>	15	0,24
1862	<i>SS 20</i>	38	0,31	1912	<i>conversion</i>	10	0,24
1863	<i>référendum</i>	27	0,31	1913	<i>nationaliser</i>	29	0,24
1864	<i>nommer</i>	20	0,31	1914	<i>Libye</i>	41	0,23
1865	<i>département</i>	62	0,30	1915	<i>libyen</i>	25	0,23
1866	<i>alliance</i>	42	0,30	1916	<i>compensatoire</i>	13	0,23
1867	<i>électeur</i>	11	0,30	1917	<i>Martinique</i>	12	0,22
1868	<i>Habré</i>	25	0,30	1918	<i>libanais (adj.)</i>	13	0,22
1869	<i>Hissène</i>	25	0,30	1919	<i>Elkabbach</i>	14	0,22
1870	<i>socialisme</i>	21	0,30	1920	<i>Guadeloupe</i>	13	0,21
1871	<i>armement</i>	92	0,30	1921	<i>terrorisme</i>	52	0,21
1872	<i>francophonie</i>	14	0,30	1922	<i>relance</i>	32	0,20
1873	<i>Gorbatchev</i>	16	0,30	1923	<i>tonne</i>	35	0,20
1874	<i>santé</i>	16	0,30	1924	<i>Israël</i>	71	0,19
1875	<i>Versailles</i>	12	0,29	1925	<i>Iran</i>	53	0,19
1876	<i>soldat</i>	30	0,29	1926	<i>université</i>	24	0,18
1877	<i>armée</i>	89	0,29	1927	<i>impôt</i>	91	0,18
1878	<i>noir</i>	23	0,29	1928	<i>étudiant (subst.)</i>	34	0,18
1879	<i>consommation</i>	34	0,29	1929	<i>candidature</i>	10	0,18
1880	<i>francophone (adj.)</i>	28	0,29	1930	<i>délabrement</i>	12	0,17
1881	<i>otage</i>	37	0,29	1931	<i>Tchad</i>	114	0,16
1882	<i>cinquième</i>	36	0,29	1932	<i>Bretagne</i>	22	0,16
1883	<i>arme</i>	109	0,29	1933	<i>Alsace</i>	17	0,14
1884	<i>investissement</i>	54	0,28	1934	<i>chaîne</i>	76	0,14
1885	<i>protectionnisme</i>	10	0,28	1935	<i>Alexandre</i>	12	0,12
1886	<i>montant</i>	17	0,28	1936	<i>lycéen (subst.)</i>	11	0,11

Une idée n'a pas besoin d'être massivement répétée pour s'imposer ; il suffit qu'elle revienne régulièrement, en toutes occasions, et dans de multiples formulations voisines. Ces mots réguliers tissent alors une présence, rendent familier un message dont l'essentiel est maintenant apparent.

### **Le vocabulaire circonstanciel.**

A l'inverse du vocabulaire habituel, les mots localisés sont signalés par des indices de répartition faibles. Ces mots peuvent survenir en une circonstance spéciale, connaître une fréquence d'emploi élevée pendant quelque temps, puis disparaître. Autrement dit leur emploi est lié à des circonstances particulières et ils ne font pas partie du vocabulaire habituel de l'auteur. Le tableau 3 présente, pour le premier septennat Mitterrand, les cent mots les plus irréguliers de fréquence égale ou supérieure à dix.

Les tableaux 1 et 3 permettent de vérifier certaines intuitions et notamment le fait que les noms propres - même très fréquents - ont souvent une répartition irrégulière. L'emploi des noms de lieux ou de personnes est généralement conditionné par la situation d'énonciation et ces mots contribuent ainsi à "particulariser" le discours. Ainsi Gorbatchev figure dans la liste - parce qu'il n'arrive au pouvoir qu'à la fin 1985 - et non point Reagan puisque celui-ci est président des Etats-Unis pendant tout le septennat<sup>10</sup>. On notera cependant que des noms très fréquents peuvent avoir une répartition irrégulière : en témoigne *Europe* (indice : 0,34) par opposition à *France* et à *Français* (0,54 et 0,51)... En fait, comme nous le signalons par ailleurs, F. Mitterrand semble découvrir l'Europe tardivement et ce fait est attesté par une répartition irrégulière. De même en bas de la liste, on trouve les noms de régions ou de départements français - *l'Alsace*, la *Bretagne*, la *Guadeloupe* et la *Martinique* - dont le président ne parle qu'à l'occasion de ses déplacements officiels dans ces territoires.

Des thèmes ponctuels se dessinent. Par exemple, *La Nouvelle-Calédonie* qui a le même indice de répartition que *Pisani* (0,28) très voisin de celui de *Canaques* (0,26). Ou encore, avec une répartition pratiquement semblable : *Hissein Habré*, le colonel *Kadhafi*, la *Libye* et les *Libyens* qui sont employés à propos de la guerre du Tchad et uniquement à ce propos... Dans le même registre international, on trouve le *Proche Orient* avec les *Palestiniens*, les *pays arabes*,

---

<sup>10</sup> M. Reagan est cité 26 fois (indice de répartition 0,5) ; M. Gorbatchev 16 fois avec un indice de 0,3.

le *Liban* et *Israël*, le *terrorisme* et les *otages*. Ou encore : l'*Union Soviétique*, les *superpuissances*, les *fusées*, les missiles *SS 20* et les *pershing* de l'*alliance (atlantique)*.

On notera tout en bas de la liste des substantifs : *lycéen*, *étudiant*, *université*. Leur indice très faible indique qu'ils sont purement conjoncturels au moins dans les thèmes développés publiquement par le président puisqu'il n'en parle qu'à propos de la loi Devaquet (décembre 1986) et de l'inauguration du salon de l'étudiant (mars 1987). A l'opposé, on notera que la *recherche scientifique* est l'un des syntagmes les plus stables du discours présidentiel (tableau 2). Si l'on s'en était tenu à la seule fréquence, on aurait pu penser que l'*université* (24 occurrences) et la *recherche* (29 occurrences dont 11 "recherche scientifique") tenaient des places à peu près égales dans les préoccupations du président. En fait, l'importance de la première tient aux circonstances (les manifestations de l'automne 1986) alors que celle de la seconde résulte d'un emploi habituel dans le discours présidentiel<sup>11</sup>.

D'autres exemples peuvent confirmer cette thèse. On remarque dans le tableau 3 des mots attachés certaines époques du septennat et discrets voire absents des autres. Les thèmes de la première année disparaissent très rapidement : le *socialisme*, la *croissance*, la *relance de la consommation* ou le verbe *nationaliser* qui est le verbe le plus localisé avec un indice de 0,24. Ils sont relayés, à partir du printemps 1983, par la *sidérurgie*, la *Lorraine*, les *tonnes d'acier*, le *délabrement de l'industrie*, sa *conversion* ou encore les *montants compensatoires*, les *impôts*, etc. Outre le vocabulaire propre à l'affaire tchadienne, la troisième période du septennat est caractérisée par quatre groupes de mots localisés : les *électeurs* et le *rassemblement national* (les élections de mars 1986 sont proches), un entretien avec M. Elkabbach en décembre 1985 qui porte principalement sur la visite du général Jaruzelski à Paris (*Pologne* et *Varsovie*), les visites dans les départements notamment d'*outre-mer* (ce qui est l'occasion d'évoquer leur relations avec la *métropole*) et enfin la création de nouvelles *chaînes de télévision*. Enfin la cohabitation est la plus fertile en mots localisés ce qui est logique puisqu'elle est très décalée par rapport au reste du vocabulaire du septennat. C'est notamment le cas de l'entretien avec Philippe Alexandre sur les *manifestations des étudiants et lycéens* à propos de la *réforme de l'université*, thème dont nous avons dit la singularité. Notons également les interventions sur les musées du *Louvre* et d'*Orsay* ou sur la *défense de la langue française* et de la *francophonie* qui

---

<sup>11</sup> Beaucoup d'autres caractéristiques du vocabulaire de F. Mitterrand confirment cette intuition : il parle très volontiers de la *science*, de la *technologie* ou de l'*informatique* et avec plus de réticence de l'*enseignement*, même *supérieur*. Il faut pour cela des circonstances politiques exceptionnelles.

suggèrent bien la stratégie de F. Mitterrand durant cette dernière période : laisser planer le doute sur sa candidature et apparaître au-dessus des contingences de la politique française...

Nous pensons que cette rapide présentation aura fait pressentir l'intérêt de la notion de répartition : elle ne se contente pas de pondérer la fréquence, elle apporte une dimension supplémentaire et révèle la singularité de certains mots. On peut lui trouver encore un autre avantage en étendant la notion à l'ensemble du vocabulaire.

Au lieu de prendre en compte chacun des mots individuellement, il est possible de considérer les catégories grammaticales et d'affecter à chacune d'elles un indice de répartition calculé de la même manière. Les résultats de ce calcul sont donnés dans le tableau 4.

Tableau 4 Indice moyen de répartition des catégories grammaticales dans l'ensemble du corpus (classement par indices décroissants).

Catégorie	N*	Fréquence moyenne	R	écart type
Prépositions	50	1182	0,64	0,03
Déterminants	48	1083	0,62	0,08
Adverbes	127	198	0,60	0,03
Verbes	456	150	0,59	0,08
Pronoms	47	976	0,57	0,04
Adjectifs	274	45	0,54	0,08
Substantifs	852	48	0,50	0,09
Noms propres	79	55	0,43	0,13

\* nombre de mots pris en compte dans le calcul

Suivant les indications du calcul, il est possible de distinguer plusieurs groupes dont le comportement diffère assez sensiblement. Le coeur du vocabulaire habituel semble bien formé de quelques outils que la langue impose plus ou moins : prépositions, articles, adverbes, pronoms auxquels il faut ajouter quelques verbes usuels (notamment *être, avoir, dire, faire...*). Dans l'ensemble ces catégories sont homogènes : on observe une dispersion assez faible autour

de la moyenne<sup>12</sup>. En revanche, les substantifs, les adjectifs et surtout les noms propres connaissent une répartition plus irrégulière et leur comportement paraît divers. Pour un petit nombre d'entre eux, on obtient des indices élevés ce qui les rattache au vocabulaire habituel et signale des tics de langage, des habitudes de pensée, des préoccupations durables. Au contraire, la plupart des adjectifs et des noms appartiennent au vocabulaire circonstanciel comme l'indiquent la moyenne et la forte dispersion autour de celle-ci.

Enfin on peut calculer une moyenne d'ensemble qui pourrait être un bon indicateur de la propension plus ou moins grande d'un auteur à spécialiser son discours. Ainsi sur le corpus Mitterrand, on obtient un indice de répartition moyen de 0,58 contre 0,59 pour ses interventions dans son débat avec J. Chirac. Cette stabilité est-elle accidentelle ou bien peut-on considérer que cette moyenne représente une caractéristique de son style ? Avant de conclure, un nombre plus important de mesures devront être effectuées notamment sur d'autres hommes politiques.

Nous espérons que cette courte note aura convaincu le lecteur de l'intérêt de cette notion de répartition pour les études de lexicographie. L'analyse des textes du premier mandat présidentiel de F. Mitterrand que nous avons résumée ici montre que l'indice de répartition n'est pas une simple pondération de la fréquence. Il apporte des informations supplémentaires, trace des lignes de force suggestives, impose des rapprochements inattendus qui se révèle souvent féconds parce qu'il permet au lexicographe d'appréhender certains aspects de la structure des textes qu'il étudie.

---

<sup>12</sup> La catégorie des "déterminants" est plus hétérogène. Il faudrait distinguer les articles et adjectifs indéfinis d'un côté, les numéraux et les cardinaux de l'autre. Les premiers ont une répartition régulière et s'écartent peu de la moyenne alors que les seconds sont plutôt irréguliers et plus dispersés (voir par exemple dans le tableau 1 le cas de *cent, mille, neuf*, etc).