

# Les mots démocrates et les expressions républicaines: analyse statistique des discours de B. Obama et de J. McCain

Jacques Savoy  
Institut d'informatique, Université de Neuchâtel

Le processus électoral américain arrive dans sa dernière ligne droite après les conventions nationales des deux grands partis. Si la télévision analyse les tendances les plus récentes et possibles développements en retenant une simple phrase de l'un des candidats, nous désirons dresser une image plus complète en tenant compte de l'ensemble des discours des deux candidats. Grâce à Internet, nous avons pu récupérer tous les discours prononcés par Barack Obama depuis sa déclaration de candidature (10 février 2007) ainsi que ceux de John McCain (depuis le 25 avril 2007). Cet ensemble composé de 182 discours (73 ou 40 % pour McCain, 109 ou 60 % pour Obama) comprend un total de 473 844 mots (172 349 ou 36,4 % pour McCain, 301 495 ou 63,6 % pour Obama) qui ont été analysés par nos outils informatiques de statistique lexicale. Sur cette base, nous pouvons obtenir une image plus complète des grands thèmes de cette élection d'une part et, d'autre part, des caractéristiques propres à chacun des candidats.

Les mots les plus fréquents dans l'ensemble des discours ne s'avèrent, à première vue, pas très intéressants. On y retrouve des déterminants ("the" étant la forme la plus fréquente), des prépositions ("to", "of") des conjonctions ("and", "but") ou des verbes auxiliaires ("is", "have"). Après ces mots grammaticaux ayant essentiellement une fonction de liaison, nous retrouvons les indices plus précis dénotant le contenu de politique américaine avec les formes "America(n)" (rang 27 et 39), "year" (42), "people" (43), "country" (48), "time" (53), ou "president" (61). Les thèmes récurrents traversant toute la campagne apparaissent également comme "war" (65), "Iraq" (85), "change" (91), "health" (93), "energy" (103), "economy" (115) ou "security" (126). Etant en présence de discours électoraux, les formes "tax" (84), "jobs" (107) ou "cost" (123) connaissent également une forte fréquence d'occurrence.

Nous avons ensuite comparé cette liste des mots les plus fréquents dans les discours électoraux avec celle obtenue par analyse de l'usage quotidien de la langue américaine. Cette information est disponible dans le corpus de Brown [FRA 82], du nom de l'université américaine qui, au début des années 1960, a récolté de manière systématique des textes provenant de diverses sources (nouvelles sportives ou économiques, roman, livres d'éducation, etc.). Lors de cette comparaison, on constate que les politiciens ont tendance à employer plus souvent le "je" (soit les formes "I" (rang 9 vs. 13 dans le corpus de Brown) ou "me" (82 vs. plus de 10 000)) ou le "nous" ("our" (11 vs. 13) ou "we" (8 vs. 24)). A l'inverse, certaines formes apparaissent plus rarement comme le "he" (rang 54 vs. 7 dans le corpus de Brown) ou "she" (205 vs. 19) de même que "woman" (694 vs. 199). La politique américaine serait-elle sexiste, étant une affaire d'hommes (dans laquelle Hillary n'aurait pas vraiment eu sa chance ?). Par rapport à la campagne présidentielle française de 2007 [SAV 07], nous constatons, pour les deux pays, un suremploi marqué du "je", du "nous" et de la désignation du pays. Par contre, le terme "politique", fréquent en France, ne l'est pas vraiment dans le discours américain qui utilisera de manière plus significative le mot pays ("country") ou peuple ("people").

Le recours plus fréquent à un ensemble donné de mots plutôt qu'à un autre ouvre d'autres perspectives. On peut ainsi vérifier quelles sont les thèmes se rattachant plus à l'un des candidats ainsi que la manière qu'il a de les aborder. Durant l'année 2008, le mot le plus spécifique du discours de John McCain s'avère être "Obama" tandis que pour Barack Obama il s'agit de "McCain". Ce premier fait relève deux aspects. D'abord les candidats utilisent la première personne pour expliquer leur point de vue (comme, par exemple, "as President I will", "If I'm elected President"). De plus, ils présentent leurs arguments en les comparant directement et explicitement avec ceux de leur adversaire, d'où la nécessité de désigner clairement ce dernier.

Si on consulte les autres formes plus spécifiques au discours de John McCain, on y retrouve "federal", "Canada", "judicial", "reform", "court", "Arizona", "business". On se souviendra que ce candidat est également sénateur de l'Arizona et qu'il a fait un voyage au Canada. Ce pays constitue également l'un des partenaires importants des Etats-Unis. De plus, on constate également le souci de faire des réformes, en particulier dans le système judiciaire. Le terme "criminal", représentatif du discours républicain permet de mettre en lumière les réformes souhaitées. Au niveau économique, le candidat adopte le vocabulaire des entreprises ("business" mais aussi "market" ou "dollar"). Parmi les termes sous-employés par le candidat du parti républicain, on y retrouve les termes "change", "kids", "politics", "Washington", "Bush", "not", "why" ou "because". Visiblement, le candidat connaît des difficultés à se démarquer de l'image de l'administration Bush et hésite à y faire référence comme aux cercles politiques de la capitale. Enfin, il use peu de la négation ou des formes annonçant une explication ("why" ou "because"), contrairement au candidat démocrate.

Pour ce dernier, les mots sur-employés sont "Bush", "families", "working", "jobs", "billion", "corporation", "election", "she", "patriotism" ou "Kennedy". Clairement le candidat désire présenter une alternative à la ligne suivie par Bush et préfère parler d'économie en terme de "places d'emploi" ou en désignant clairement des grandes entreprises. Contrairement à son adversaire, Barack Obama cherche à expliquer ("because") et tient un discours sur la famille ou l'énergie renouvelable. Une place, certes relativement réduite, est faite aux femmes ("she") contrairement aux discours du candidat républicain. Enfin, il ne néglige pas, depuis quelques mois surtout, à noter son attachement à la nation ("patriotism"). Parmi les mots que Barack Obama n'utilise que fort peu, on retrouve "islamic", "property", "gun", "law" ou "government". Le candidat hésite à discuter du 2e amendement (libre port d'armes à feu) ou d'indiquer qu'il n'est pas musulman.

Enfin, nous pouvons examiner les séquences les plus fréquentes et composées de deux ou trois mots. Dans le champ démocrate, on préfère les séquences "that is why", "it is time", "the American people" tandis que du côté républicain, on adopte facilement les formes "the United States", "we need to", "in the world". Si l'on cherche des séquences portant une sémantique plus précise, le discours de Barack Obama comprend l'usage récurrent des formes "universal health care", "million new jobs" ou "world class education" tandis que John McCain introduit volontiers "(our) health care system", "small business owner(s)" ou "(our) dependance (on) foreign oil".

Une campagne électorale ne se fait pas en un jour et les candidats doivent adapter leur discours au fil des mois, selon les événements extérieurs, l'audience et les préoccupations particulières de cette dernière. Ainsi, on ne recourt pas au même vocabulaire devant une assemblée d'hommes d'affaires ou d'enseignants. En découpant par mois les discours de John McCain durant l'année 2008, nous voyons apparaître comme mots spécifiques "deregulation" et "WTO" en janvier, "bubble" (bulle spéculative), "submarine" (son père était officier sous-marinier) en mars ou

“judicial”, “court” et “judges” en mai. En juillet, John McCain employait de manière plus abondante les termes “sheriff”, “criminal” ou “jobs” et, en août, “veteran”, “(American) Legion” et “(health) care” (pour les anciens combattants). Mais la dynamique a des effets également chez le candidat démocrate où les formes “Berlin”, “(men and) women” ou “partnership” sont caractéristiques des discours tenus en juillet tandis que les mots “Joe Biden”, “oil”, “energy” ou “renewable” marquent ses interventions du mois d'août.

Au niveau de la campagne en général, les deux derniers mois peuvent se caractériser, pour le mois de juillet, par l'incursion des termes “Berlin”, “cyber” et “NAACP” (organisation pour la promotion des droits civils des minorités) de même que la disparition des mots “Israel”, “health” et “war”. Le mois d'août a vu une surreprésentation des mots “veteran”, “Joe Biden”, “oil” et “urban” et comme corolaire l'abandon des formes “trade”, “world” et “market”. La crise de l'énergie et les problèmes des villes tiennent une place plus significative dans la campagne. Finalement, on devrait s'attendre à une remontée du “she” dans les fréquences d'occurrences ainsi que l'apparition du nom “Sarah Palin”, du moins dans le discours de John McCain.

### Référence

[FRA 82] FRANCIS W. N., KUČERA H.: *Frequency Analysis of English Usage, Lexicon and Grammar*. Houghton Mifflin, Boston, 1982.

[SAV 07] SAVOY J.: *Sous l'éclairage de la statistique lexicale, les mots de Ségolène et de Nicolas*. Le Temps, 22 juin 2007.

Brown			Politique		
rang	mot	fréquence	rang	mot	fréquence
1	the	6.90%	1	the	4.89%
2	be	3.86%	2	and	3.83%
3	of	3.59%	3	to	3.35%
4	and	2.85%	4	of	2.78%
5	to	2.58%	5	that	2.14%
6	a	2.28%	6	a	2.01%
7	in	2.06%	7	in	1.95%
8	he	1.92%	8	we	1.77%
9	have	1.23%	9	is	1.49%
10	it	1.08%	10	i	1.39%
11	for	0.89%	11	for	1.17%
12	i	0.83%	12	our	1.16%
13	they	0.82%	13	not	1.03%
14	with	0.72%	14	it	1.01%
15	not	0.69%	15	have	0.89%
16	that	0.64%	16	will	0.82%
17	on	0.61%	17	are	0.79%
18	she	0.60%	18	this	0.74%
19	as	0.59%	19	on	0.58%
20	at	0.53%	20	as	0.57%

