DISCOURS D'INVESTITURE DE LA CINQUIÈME : ESSAI DE TEXT MINING

LE 24 JUILLET 2011 ***MARIE COUSSIN***

Les six discours d'investiture des Présidents de la République sont passés à la moulinette: quels présidents étaient les plus bavards? Quels mots revenaient le plus dans leurs discours? Lequel parle le plus souvent de lui?

Billet initialement publié sur le **datablog d'OWNI**

Pour ce test, j'ai choisi les discours prononcés par les présidents de la Vème République lors de leur (première) investiture afin de pouvoir comparer des textes énoncés dans un même contexte.
Ils sont tous accessibles et exportables en PDF (faut pas trop en demander non plus…) sur le site de **l'Elysée**.
Pour lire le détail des textes :
Discours de Charles De Gaulle, 8 janvier 1959
Discours de Georges Pompidou, 20 juin 1969
Discours de Valéry Giscard d'Estaing, 27 mai 1974
Discours de François Mitterand, 21 mai 1981
Discours de Jacques Chirac, 17 mai 1995
Discours de Nicolas Sarkozy, 16 mai 2007

De Gaulle et Sarkozy, les plus bavards
Premier élément à comparer : la longueur des discours, dont la différence saute aux yeux une fois le nombre de mots extrait grâce au logiciel de traitement de texte.

Qui parle le plus, et de quoi ?
Pour rentrer un peu plus dans le détail des textes, j'ai essayé **OpenCalais** qui crée des metadonnées sur les noms d'entités, faits et évènements repérés dans un texte.
Dans ce cas précis de discours d'investiture, et avec sa version démo, OpenCalais n'apportait pas vraiment de valeur ajoutée : les lieux, personnes et institutions citées dans un discours d'investiture sont un peu toujours les mêmes.
Voilà ce qu'OpenCalais extrait pour le discours du Général De Gaulle :

J'ai ensuite testé le logiciel **Tropes**, pingué par notre collègue data Guénaël Pépin (**@Reguen** sur Twitter).

Fonctionnant en deux versions, anglaise et française, il offre quelques fonctionnalités d'analyses assez fines des discours : style d'énonciation, scénario de construction,

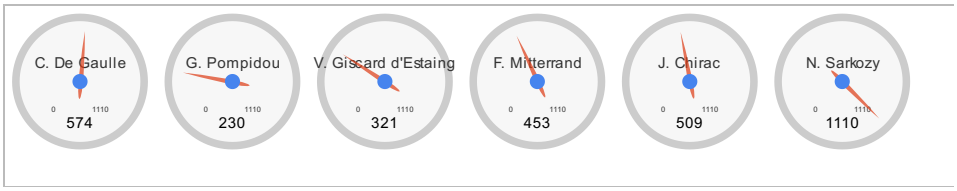
verbes/adjectifs/substantifs les plus prononcés, etc.
Tropes permet aussi d'extraire des graphs (mais ils sont loin d'être esthétiquement exploitables) et surtout de faire facilement des recherches sur certains termes.
Pour donner une idée, quelques éléments d'analyse sémantique produits par Tropes :

(All)	(All)	(All)
Président	Style	Références utilisées
N.Sarkozy	Style plutôt argumentatif	nation france exigence français respect monde h
J.Chirac	Style plutôt argumentatif	nation france français président
F. Mitterrand	Style plutôt argumentatif	nation français france accalmie homme
V. Giscard d'Estaing	Style plutôt énonciatif	monsieur français changement président chef_d
G. Pompidou	Style plutôt argumentatif	nation chef_d_état france

Goog

J'ai utilisé Tropes pour isoler le nombre de mots prononcés par chacun sur trois champs lexicaux forcément utilisés par les orateurs :
– l'utilisation de la première personne ;
– la référence à la France et aux Français ;
– la référence à la République.

Ces données sont à mettre en relation avec la longueur respective de chaque discours (graphique n°1), visualisé de manière différente grâce aux widgets de Google Chart…



Google Docs

Les discours dans les nuages
Dernier outil utilisé, afin d'avoir une vue d'ensemble du ton du discours, ManyEyes et son générateur de nuages de mots.
Si elle ne peut clairement pas être considérée comme une analyse scientifique précise, cette technique permet d'associer visualisation esthétique, rapidité d'exécution et éléments d'analyse.
Pour plus de lisibilité, j'ai limité le nombre de mots à 80 et ai enlevé les mots non porteurs de sens dans cette situation ("qu'il", "ceux", "le", "la", etc.)
Général De Gaulle :

Georges Pompidou :

Valéry Giscard d'Estaing :

François Mitterand :

Jacques Chirac :

Nicolas Sarkzoy

Ces visualisations permettent de mettre en valeur certaines caractéristiques :
- chaque Président a un ou plusieurs thèmes qui ressortent de leur discours : **"communauté "** pour De Gaulle, **"République"** pour Pompidou, **"changement"** pour VGE, **"tous"** et **"Français"** pour Mitterand, **"plus"** pour Chirac et enfin **"exigence"** pour Nicolas Sarkozy.
- le discours de De Gaulle rappelle clairement le contexte historique dans lequel il a été prononcé : il y parle de *"Dieu"* et fait référence à *"l'Afrique"*, aux *"africains"* et la *"Libération"*. Même effet pour Georges Pompidou : les références à *"De Gaulle"*, au *"Général"* sont très présentes, tout comme le champ lexical de la responsabilité et du sérieux *"confiance", "autorité", "institutions", "dépositaire", "gravité", "charge", "devoir".*
- le discours de Giscard d'Estaing apparaît plus centré sur l'humain *"femmes", "jeunes", "hommes"* et les réformes qu'il entend mener *"nouvelle", "politique", "conduirai", "associera", "droits", "suffrages".*
- Avec des termes comme *"peuple", "communauté", "majorité"* mais aussi *"monde", "pouvoirs" "haute", "ambition", "véritable", 'millions"* la narration de François Mitterand est davantage axé vers une vision de la France, de sa place dans le monde. On note également la référence à *"Jaurès"*.
- Le discours de Jacques Chirac met l'accent sur la *"Nation"*, la *"République"* et le champ lexical de la volonté : *"ferai", "changement", "charge", "victoire", "commence", "contre"*.
- Enfin, celui de Nicolas Sarkzoy utilise beaucoup d'adverbes *"toujours", "jamais", "aussi"* et est marqué par la notion *"d'exigence"*, de vouloir *"veut"*.
Conclusion : sur des discours prononcés dans une situation similaire et relativement contrainte (parler de la Nation, de la République, des perspectives, etc.), ces outils simples de dataviz et d'analyse sémantique permettent de mettre en lumière des spécificités liées à chaque personnalité politique.

LAM SON
le 24 juillet 2011 - 15:45 • SIGNALER UN ABUS - PERMALINK

Sans normalisation par la longueur du discourt, le graphique de récurrence de certains termes est peu utilisable.

VOUS AIMEZ 0 VOUS N'AIMEZ PAS 0 LUI RÉPONDRE

MICHEL
le 25 juillet 2011 - 10:07 • SIGNALER UN ABUS - PERMALINK

@lam. Exact. Cela dit la démarche est intéressante. Reste la rigueur méthodologique a peaufiner.

VOUS AIMEZ 0 VOUS N'AIMEZ PAS 0 LUI RÉPONDRE

BERTRAND
le 25 juillet 2011 - 14:50 • SIGNALER UN ABUS - PERMALINK

Pour De Gaulle, ce qui ressort, c'est la "Communauté" (à savoir le projet de Commonwealth lancé par la France après la décolonisation) avec un grand "C" et non la "communauté", qui est aussi présente dans le discours de Mitterrand sous son sens commun.

VOUS AIMEZ 0 VOUS N'AIMEZ PAS 0 LUI RÉPONDRE

BOREL
le 10 octobre 2012 - 17:22 • SIGNALER UN ABUS - PERMALINK

j'apprécie votre analyse et j'aimerais vous contacter pour que vous puissez m'aider dans le même sens. aidez moi. merci

VOUS AIMEZ 0 VOUS N'AIMEZ PAS 0 LUI RÉPONDRE

2 pings

Sélection de liens (weekly) « J'ai rarement tort … le 31 juillet 2011 - 9:31

[...] Discours d'investiture de la cinquième : essai de text mining » OWNI, News, Augmented [...]

Open data: Challenge accepted! » OWNI.eu, News, Augmented le 4 août 2011 - 17:45

[...] Simon Rogers from the Guardian underlines, tons of free software exists that anyone can use to easily generate visualizations and data: Google Charts, Google Fusion Tables, Many Eyes, and Timetric. However, most citizens have neither [...]