

Analyse statistique des transcriptions des meetings du FOMC (Federal Open Market Committee)

Etienne Le Naour

ENSAE & CREST

Maître de stage : Alessandro Riboni

Stage de fin d'études

Table des matières

- 1 Introduction
- 2 Exploration des données et premières analyses
- 3 Analyse de la polarité
- 4 Latent Dirichlet Allocation (LDA)
- 5 Conclusion

- Qu'est-ce que le FOMC et quels sont ses objectifs?
 - ➊ Réalise des opérations d'open market dans le but de fixer le taux d'intérêt directeur
 - ➋ Est le principal instrument de la politique monétaire américaine
- Quel est le déroulé des réunions du FOMC
 - ➊ Résumé du chair
 - ➋ Premier tour de table
 - ➌ Deuxième tour de table
 - ➍ Vote

- Problématique : mesurer l'impact du chair sur le déroulement des réunions

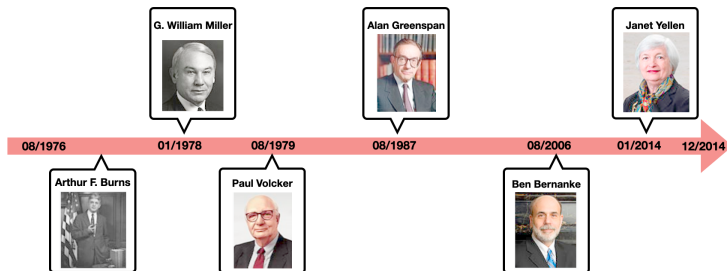


Figure: Frise chronologique des différents chairs sur le déroulé des **319** réunions étudiées

Construction de la base de données (1/3)

Différentes phases dans la construction

- ① Récupérer les transcriptions
- ② Transformer les .pdf en .txt
- ③ Nettoyer les fichiers
- ④ Associer chaque phrase à un interlocuteur
- ⑤ Construire la base de données

Construction de la bases de données (2/3)

January 24–25, 2012 1 of 314

Meeting of the Federal Open Market Committee on
January 24–25, 2012

A meeting of the Federal Open Market Committee was held in the offices of the Board of Governors in Washington, D.C., on Tuesday, January 24, 2012, at 10:00 a.m., and continued on Wednesday, January 25, 2012, at 8:30 a.m. Those present were the following:

Ben Bernanke, Chairman
William C. Dudley, Vice Chairman
Elizabeth Duke
Jeffrey M. Lacker
Dennis P. Lockhart
Sandra Pianalto
Sarah Bloom Raskin
Daniel K. Tarullo
John C. Williams
Janet L. Yellen
James Bullard, Christine Cumming, Charles L. Evans, Esther L. George, and Eric
Rosengren, Alternate Members of the Federal Open Market Committee
Richard W. Fisher, Narayana Kocherlakota, and Charles I. Plosser, Presidents of the
Federal Reserve Banks of Dallas, Minneapolis, and Philadelphia, respectively
William B. English, Secretary and Economist
Deborah J. Danker, Deputy Secretary
Matthew M. Luecke, Assistant Secretary
David W. Skidmore, Assistant Secretary
Michelle A. Smith, Assistant Secretary
Scott G. Alvarez, General Counsel
Thomas C. Baxter, Deputy General Counsel
Steven B. Kamin, Economist
David W. Wilcox, Economist
David Altig, Thomas A. Connors, Michael P. Leahy, William Nelson, Simon Potter,
David Reischneider, Glenn D. Rudebusch, and William Wascher, Associate Economists
Brian Sack, Manager, System Open Market Account
Michael S. Gibson, Director, Division of Banking Supervision and Regulation, Board of
Governors
Nellie Liang, Director, Office of Financial Stability Policy and Research, Board of
Governors

January 24–25, 2012 2 of 314

Jon W. Faust and Andrew T. Levin, Special Advisors to the Board, Office of Board
Members, Board of Governors
James A. Clouse, Deputy Director, Division of Monetary Affairs, Board of Governors
Linda Robertson, Assistant to the Board, Office of Board Members, Board of Governors
Daniel E. Sichel, Senior Associate Director, Division of Research and Statistics, Board of
Governors

Figure: Aperçu du format initial d'une transcription

Aperçu de la base de données (3/3)

- Intérêt pour la temporalité et le contenu

	Date		interlocutor_name	statement_size	statement	statement_number
80272	1995-03-28		CHAIRMAN GREENSPAN	9	would somebody like to move approval of the mi...	statement_0
80273	1995-03-28	VICE	CHAIRMAN MCDONOUGH	3	somove	statement_1
80274	1995-03-28		CHAIRMAN GREENSPAN	109	without objection the committee has received a...	statement_2
80275	1995-03-28		MS MINEHAN	2	so move	statement_3
80276	1995-03-28		CHAIRMAN GREENSPAN	25	without objection i turn to the vice chairman ...	statement_4
80277	1995-03-28	VICE	CHAIRMAN MCDONOUGH	110	thankyoumrchairmanerniepatrikiswhohasbeenservi...	statement_5
80278	1995-03-28		CHAIRMAN GREENSPAN	7	would somebody like to move that nomination	statement_6
80279	1995-03-28		MR KELLEY	2	move it	statement_7
80280	1995-03-28		CHAIRMAN GREENSPAN	4	would somebody second it	statement_8
80281	1995-03-28		MS MINEHAN	1	second	statement_9
80282	1995-03-28		CHAIRMAN GREENSPAN	27	without objection next we move to the election...	statement_10
80283	1995-03-28		MR MOSKOW	26	mr chairman i would like to nominate william c...	statement_11
80284	1995-03-28		CHAIRMAN GREENSPAN	12	is there discussion would somebody like to mov...	statement_12
80285	1995-03-28		CHAIRMAN GREENSPAN	132	moved and seconded and approved without object...	statement_13
80286	1995-03-28		MR HOENIG	78	mr chairman since it is proposed to increase t...	statement_14
80287	1995-03-28		CHAIRMAN GREENSPAN	95	the reason for the proposal to increase staff ...	statement_15
80288	1995-03-28		MR KOHN	10	president jordanssuggestion was to go from to...	statement_16
80289	1995-03-28		CHAIRMAN GREENSPAN	8	how much higher do you want to go	statement_17
80290	1995-03-28		MR JORDAN	18	i am not sure why it is helpful to have access...	statement_18
80291	1995-03-28		CHAIRMAN GREENSPAN	21	the basic presumption is that there are fewer ...	statement_19

Figure: Aperçu des 20 premières lignes de la base de données

Nombre de mots prononcés par réunion (1/2)

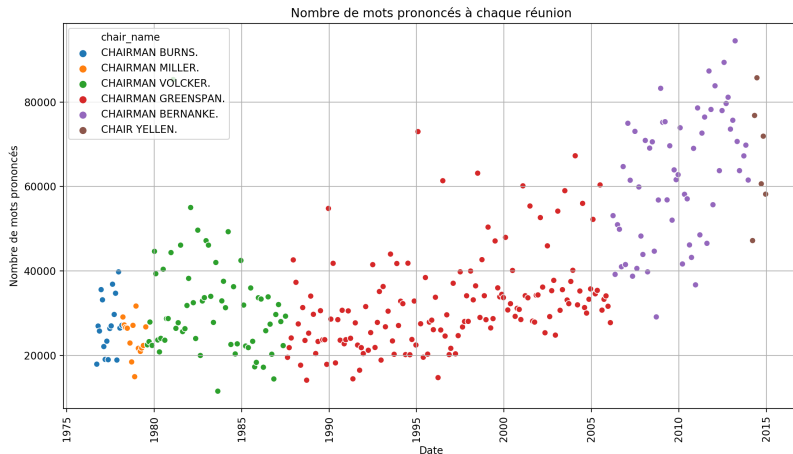


Figure: Nombre total de mots prononcés à chaque réunion pour la période 1976-2015

Nombre de mots prononcés par réunion (2/2)

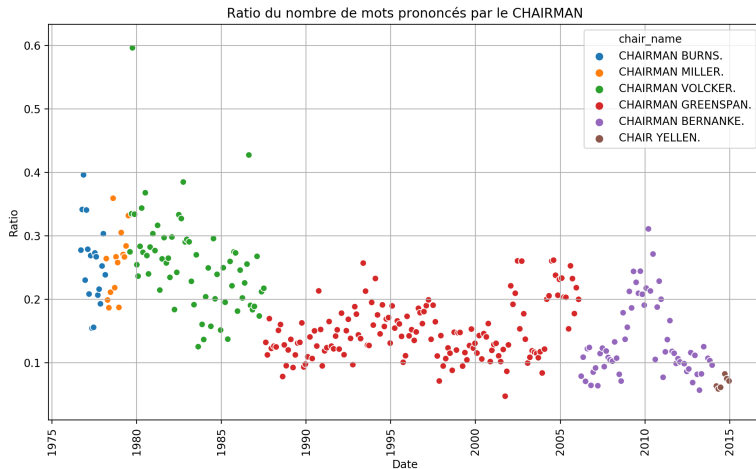


Figure: Nombre de mots prononcés par les chairs relativement au nombre total de mots prononcés au cours de la réunion sur la période 1976-2015

Une nécessité de labéliser les mots

- Intérêt pour le contenu des interventions
- Les dictionnaires de sentiments
 - Dictionnaire de Loughran and McDonald (vocabulaire négatif, positif, incertain etc.)
 - Dictionnaire d'Harvard (vocabulaire académique, économique, etc.)
- Construction de scores faisant sens (score de positivité, d'incertitude, d'affirmation)
- Regard sur l'évolution de ces scores dans le temps et essayer de déduire des relations causales entre la variation de ces scores et la présence d'un tel chair ou un autre

Aperçu du dictionnaire de Loughran and McDonald

A	B	C	D	E	F	G
Negative	Positive	Uncertainly	Litigious	StrongModal	WeakModal	Constraining
ABANDON	ABLE	ABEYANCE	ABOVENTION	ALWAYS	ALMOST	ABIDE
ABANDONED	ABUNDANCE	ABEYANCES	ABROGATE	BEST	APPARENTLY	ABIDING
ABANDONING	ABUNDANT	ALMOST	ABROGATES	CLEARLY	APPEARED	BOUND
ABANDONMENT	ACCLAIMED	ALTERATION	ABROGATES	DEFINITELY	APPEARING	BOUNDED
ABANDONMENT	ACCOMPLISH	ALTERATIONS	ABROGATING	DEFINITIVELY	APPEARS	COMMIT
ABANDONS	ACCOMPLISHED	AMBIGUITIES	ABROGATION	HIGHEST	CONCEIVABLE	COMMITMENT
ABDICATED	ACCOMPLISHES	AMBIGUITY	ABROGATIONS	LOWEST	COULD	COMMITMENTS
ABDICATES	ACCOMPLISHING	AMBIGUOUS	ABSOLVE	MUST	DEPEND	COMMITTS
ABDICATING	ACCOMPLISHME	ANOMALIES	ABSOLVED	NEVER	DEPENED	COMMITTED
ABDICATION	ACCOMPLISHME	ANOMALOUS	ABSOLVES	STRONGLY	DEPENDING	COMMITTING
ABDICATIONS	ACHIEVE	ANOMALOUSLY	ABSOLVING	UNAMBIGUOUS	DEPENDS	COMPEL
ABERRANT	ACHIEVED	ANOMALY	ACCESSION	UNCOMPROMIS	MAY	COMPELLED
ABERRATION	ACHIEVEMENT	ANTICIPATE	ACCESSIONS	UNDISPUTED	MAYBE	COMPELLING
ABERRATIONAL	ACHIEVEMENTS	ANTICIPATED	ACQUIREES	UNDOUBTEDLY	MIGHT	COMPELS
ABERRATIONS	ACHIEVES	ANTICIPATES	ACQUIRORS	UNEQUIVOCAL	NEARLY	COMPLY
ABETTING	ACHIEVING	ANTICIPATING	ACQUIT	UNEQUIVOCALLY	OCCASIONALLY	COMPULSION
ABNORMAL	ADEQUATELY	ANTICIPATION	ACQUITS	UNPARALLELED	PERHAPS	COMPULSORY
ABNORMALITIES	ADVANCEMENT	ANTICIPATIONS	ACQUITTAL	UNSURPASSED	POSSIBLE	CONFIN

Figure: Dictionnaire de Loughran and McDonald de polarité

Définition des scores de polarité

- $ScorePositive = \frac{\sum_{word} (1_{word \in Positive} - 1_{word \in Negative})}{\sum_{word} (1_{word \in Positive} + 1_{word \in Negative})}$
- $ScoreAffirmation = \frac{\sum_{word} (1_{word \in StrongModal} - 1_{word \in WeakModal})}{\sum_{word} (1_{word \in StrongModal} + 1_{word \in WeakModal})}$
- $ScoreIncertitude = \frac{-\sum_{word} 1_{word \in Incertitude}}{NbWord}$

- Ces scores peuvent être appliqués à une personne ou bien être globalisés à l'ensemble de la réunion
- L'évolution de ces scores dans le temps
- Possibilité de construire d'autres scores

Comment exploiter ces scores?

- On aimerait mesurer l'influence d'un chair sur la positivité (ainsi que l'incertitude et l'affirmation) des propos tenus par les autres membres de la réunion de façon causale
- La période Greenspan-Bernanke semble particulièrement intéressante
- On essaie de mesurer le fait qu'un des deux chairs soit en charge de la réunion plutôt que l'autre

Rappel sur la régression avec fixed effect

Régression avec fixed effect

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_t + \gamma_2 D2_i + \gamma_3 D3_i + \dots + \gamma_n Dn_i + u_{it}$$

où

- X_{it} désigne la variable explicative. Dans notre cas une indicatrice.
- Y_{it} désigne la variable à expliquer. Dans notre cas cette variable désignera un score propre à un membre i du comité à un moment t .
- Z_t désigne les variables de contrôles.
- β_0 désigne la constante du modèle, β_1 désigne le coefficient à la variable explicative d'intérêt, β_2 désigne le coefficient associé à la variable de contrôle.
- Les Dj_i désignent des indicatrices qui captent les effets fixes propres à un individu i .
- Les γ sont les coefficients associés à ces effets fixes.

Résultat de la régression sur le score de positivité

```

Coefficients:
(Intercept)          -0.234570    0.031910   -7.351 2.86e-13 ***
ChairCHAIRMAN GREENSPAN.  0.104378    0.018646    5.598 2.47e-08 ***
NameMR. GUYNN.        -0.024841    0.045609   -0.545 0.58606
NameMR. HOENIG.       0.122950    0.038552    3.189 0.00145 **
NameMR. KAMIN.        -0.223975    0.072975   -3.069 0.00218 **
NameMR. KOHN.         -0.190977    0.037866   -5.043 4.99e-07 ***
NameMR. KOS.          -0.086483    0.055260   -1.565 0.11774
NameMR. LACKER.       -0.041460    0.047496   -0.873 0.38281
NameMR. MADIGAN.      -0.014193    0.060059   -0.236 0.81321
NameMR. MEYER.        -0.145339    0.054944   -2.645 0.00823 **
NameMR. MOSKOW.       0.012768    0.043246    0.295 0.76784
NameMR. OLSON.        0.119981    0.061859    1.940 0.05257 .
NameMR. POOLE.        -0.080826    0.046393   -1.742 0.08162 .
NameMR. REIFSCHNEIDER. -0.203902    0.070503   -2.892 0.00387 **
NameMR. REINHART.     -0.159556    0.055920   -2.853 0.00437 **
NameMR. SACK.         -0.069195    0.071988   -0.961 0.33657
NameMR. SHEETS.       -0.121977    0.067420   -1.809 0.07057 .
NameMR. SLIFMAN.      -0.120461    0.070861   -1.700 0.08930 .
NameMR. STERN.        0.078221    0.038150    2.050 0.04046 *
NameMR. STOCKTON.     -0.132413    0.042306   -3.130 0.00177 **
NameMR. WILCOX.       -0.110330    0.058462   -1.887 0.05928 .
NameMR. WILLIAMS.     -0.085006    0.073057   -1.164 0.24475
NameMS. BIES.         -0.052723    0.058446   -0.902 0.36712
NameMS. DANKER.       0.627387    0.051711   12.133 < 2e-16 ***
NameMS. JOHNSON.     -0.038615    0.049067   -0.787 0.43139
NameMS. MINEHAN.      -0.073401    0.042952   -1.709 0.08762 .
NameMS. PIANALTO.     0.018989    0.045443    0.418 0.67609
NameMS. SMITH.        0.388638    0.083708    4.643 3.66e-06 ***
NameMS. YELLEN.       -0.101072    0.043873   -2.304 0.02134 *
NameVICE CHAIRMAN GEITHNER. 0.102270    0.059195    1.728 0.08420 .
TauxInteret          -0.003753    0.003641   -1.031 0.30289
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3236 on 1991 degrees of freedom
Multiple R-squared:  0.189,    Adjusted R-squared:  0.1767
F-statistic: 15.46 on 30 and 1991 DF,  p-value: < 2.2e-16

```

Figure: Résultat sur R de la régression présentée ci-dessus

Réflexion sur les variables de contrôles, indication sur les autres résultats et limites

Bilan de ce modèle

- Critique sur les variables de contrôles
 - Résultat pour le score d'incertitude
 - Limite de notre approche
-
- Nécessité de trouver une approche basée sur des scores non arbitraires
 - Construction de scores à partir de l'approche Latent Dirichlet Allocation (LDA)

L'approche LDA : rapide explication théorique

- Modèle d'apprentissage non supervisé basé sur un corpus de documents
- Représentation d'un document sous forme de mélange de topics
- Représentation des topics comme mélange de mot

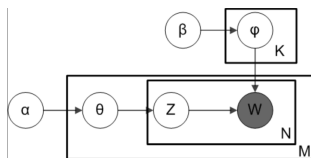


Figure: Schéma simplifié des paramètres dans le modèle LDA

- Le but est d'avoir des topics distincts les uns des autres

- Considérer qu'une réunion est un document ne marche pas, ils ne sont pas assez distinguables entre eux.
- On considère qu'un document désigne l'intégralité des mots prononcés par un membre du comité au cours de sa carrière.
- Il y a autant de document que de personnes
- On fixe de façon arbitraire le nombre de topics à 5
- On choisit parmi les 20 mots les plus représentatifs de chaque topic 5 mots pour le caractériser.

Topics en sortie de l'algorithme

- Topic 0: president market governor policy committee
 - Topic 1: percent rate growth time range
 - Topic 2: inflation policy market economy financial
 - Topic 3: forecast prices gdp market unemployment data
 - Topic 4: economy year time district people
-
- Interpréter les topics

Distribution des topics pour chaque chair

Nom du chair	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4
Chairman Burns	4%	94%	2%	0%	0%
Chairman Miller	5%	95%	0%	0%	0%
Chairman Volcker	0%	100%	0%	0%	0%
Chairman Greenspan	80%	2%	0%	1%	16%
Chairman Bernanke	48%	0%	52%	0%	0%
Chair Yellen	34%	0%	66%	0%	0%

Figure: Distribution des topics (arrondis à deux chiffres) pour chaque Chair

- Construction de scores à partir des résultats
- Regard sur la similitude entre les sujets abordés par le chair et les autres membres de la réunion
- Nombreuses pistes d'amélioration

Conclusion

- Ce que mon travail a apporté
- Un stage enrichissant personnellement