

Comment faire des sciences sociales avec des traces textuelles ?

Sylvain Parasio

Facebook profile for Sean Nicholson, March 6, 2013. Post text: "I haven't written a haiku in a while, so here's one based on a real experience today: In line at Starbucks Default iPhone ring tone sounds Pavlov was brilliant... See More"

Twitter post by SKY (@SkylaCheyenne) from Sep 19, 2016. Text: "Chaque jour, quand mon chien est triste de me voir pour l'école, je me dis : « Je fais ça pour toi. Je fais ça pour qu'on puisse vivre une belle vie. »"

Instagram comment section for the haiku post. Comment by ROBERT: "un grand merci aux agents d'ERDF qui bossaient le 24, le 25, pendant que tous leurs détecteurs baffletraient sans se soucier de leur prochain"

Le Monde.fr news website interface. Main headline: "Le gouvernement Cameron fixe ses priorités". Other headlines include "En continu", "Les blogs", "Actualités", and "Editions".

Historical newspaper page from 'l'Humanité' (Journal Socialiste). Main headline: "TRÈS VIOLENTE BATAILLE sur le front de Verdun". Other headlines include "Sur la Route DE VERDUN" and "Une habile Diversion".

Digitized book cover for 'SHAKESPEARE' by Horace Howard Furness, Ph.D., LL.D. Volume 3, Part 1. The cover features the title in large serif font and the publisher's name 'A NEW VARIORUM EDITION'.

Prolifération des « traces textuelles »

Do Anti-Immigrant Laws Shape Public Sentiment? A Study of Arizona's SB 1070 Using Twitter Data¹

René D. Flores
University of Washington



The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings

Austin C. Kozlowski,^a  Matt Taddy,^b
and James A. Evans^{a,c} 

Dissecting the Spirit of Gezi: Influence vs. Selection in the Occupy Gezi Movement

Channeling Hearts and Minds: Advocacy Organizations, Cognitive-Emotional Currents, and Public Conversation

Christopher A. Bail,^a Taylor W. Brown,^a and Marcus Mann^a

Prestige, Proximity, and Prejudice: How Google Search Terms Diffuse across the World¹

Christopher A. Bail and Taylor W. Brown
Duke University

Andreas Wimmer
Columbia University

Measuring CULTURE

.....

John Mohr • Christopher A. Bail

Margaret Frye • Jennifer C. Lena

Omar Lizardo • Terence E. McDonnell

Ann Mische • Iddo Tavory

Frederick F. Wherry

Multiplication des recherches sociologiques

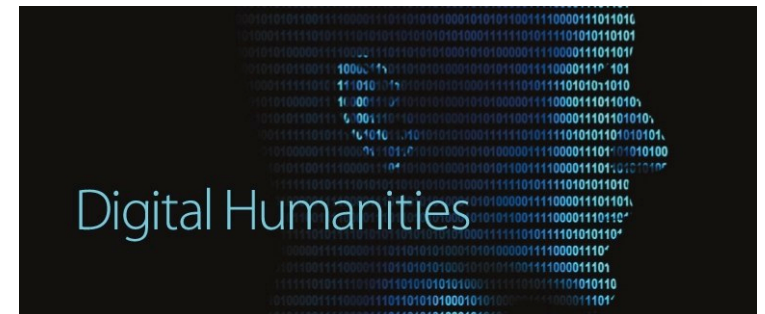
SOCIAL SCIENCE

Computational Social Science

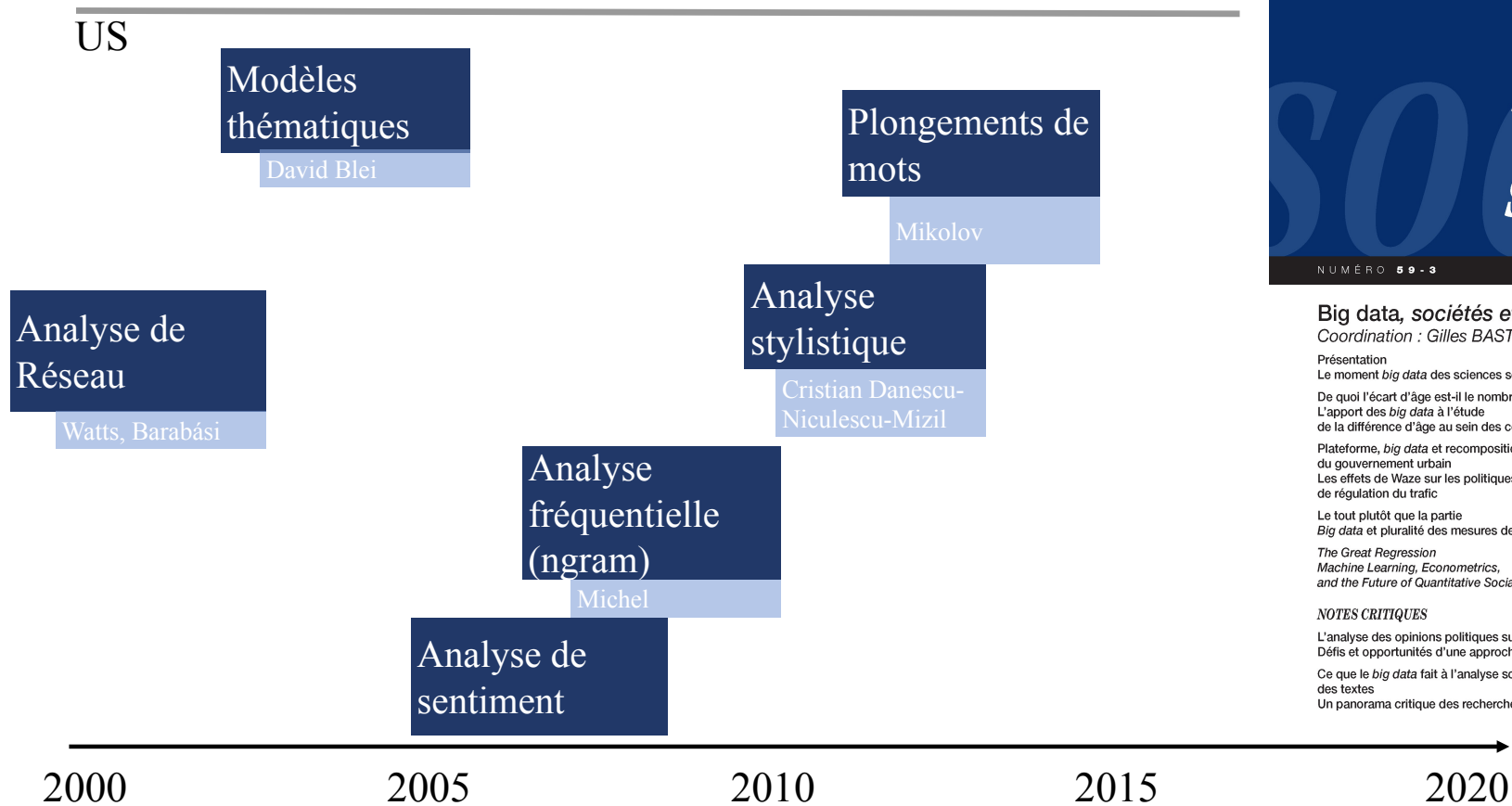
David Lazer,¹ Alex Pentland,² Lada Adamic,³ Sinan Aral,^{2,4} Albert-László Barabási,⁵
Devon Brewer,⁶ Nicholas Christakis,¹ Noshir Contractor,⁷ James Fowler,⁸ Myron Gutmann,³
Tony Jebara,⁹ Gary King,¹ Michael Macy,¹⁰ Deb Roy,² Marshall Van Alstyne^{2,11}

A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

- Être plus inductif
- Étendre l'échelle de l'analyse
- Exploiter des matériaux non sollicités
- Relier des matériaux textuels et des informations relationnelles



Promesses pour les sciences sociales



Big data, sociétés et sciences sociales

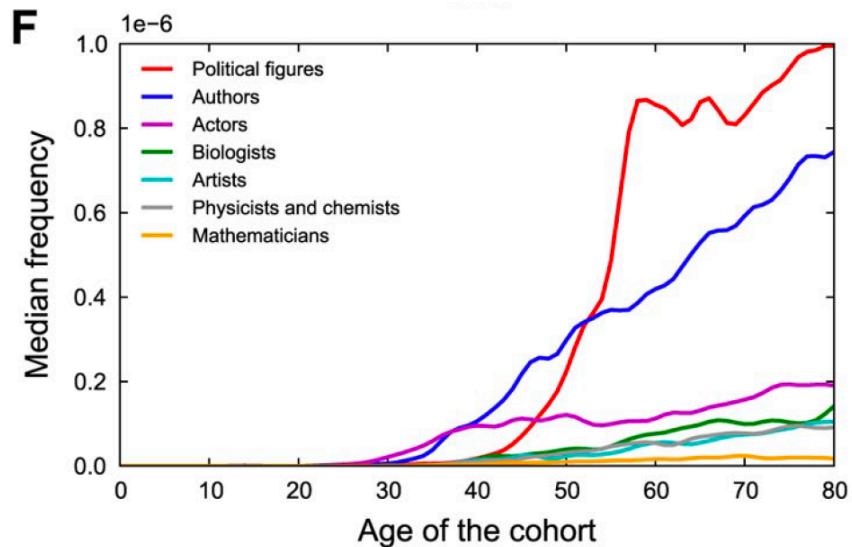
Coordination : Gilles BASTIN, Paola TUBARO

Présentation	GILLES BASTIN
Le moment <i>big data</i> des sciences sociales	PAOLA TUBARO
De quoi l'écart d'âge est-il le nombre ?	MARIE BERGSTRÖM
L'apport des <i>big data</i> à l'étude de la différence d'âge au sein des couples	
Plateforme, <i>big data</i> et recomposition du gouvernement urbain	ANTOINE COURMONT
Les effets de Waze sur les politiques de régulation du trafic	
Le tout plutôt que la partie	BAPTISTE KOTRAS
<i>Big data</i> et pluralité des mesures de l'opinion sur le web	
<i>The Great Regression</i>	JULIEN BOELAERT
<i>Machine Learning, Econometrics, and the Future of Quantitative Social Sciences</i>	ÉTIENNE OLLION
NOTES CRITIQUES	
L'analyse des opinions politiques sur Twitter	MARTA SEVERO
Défis et opportunités d'une approche multi-échelle	ROBIN LAMARCHE-PERRIN
Ce que le <i>big data</i> fait à l'analyse sociologique des textes	JEAN-PHILIPPE COINTET
Un panorama critique des recherches contemporaines	SYLVAIN PARASIE

Renouveau des techniques d'analyse textuelle

Quantitative Analysis of Culture Using Millions of Digitized Books

Jean-Baptiste Michel,^{1,2,3,4*†} Yuan Kui Shen,⁵ Aviva Presser Aiden,⁶ Adrian Veres,⁷ Matthew K. Gray,⁸ The Google Books Team,⁸ Joseph P. Pickett,⁹ Dale Hoiberg,¹⁰ Dan Clancy,⁸ Peter Norvig,⁸ Jon Orwant,⁸ Steven Pinker,⁴ Martin A. Nowak,^{1,11,12} Erez Lieberman Aiden^{1,12,13,14,15,16*†}



« Étendre les frontières de l'enquête quantitative rigoureuse à un large ensemble de phénomènes nouveaux couvrant les sciences sociales et les humanités »

Les sciences sociales concurrencées ?

Language from police body camera footage shows racial disparities in officer respect

Rob Voigt^{a,1}, Nicholas P. Camp^b, Vinodkumar Prabhakaran^c, William L. Hamilton^c, Rebecca C. Hetey^b, Camilla M. Griffiths^b, David Jurgens^c, Dan Jurafsky^{a,c}, and Jennifer L. Eberhardt^{b,1}

EXAMPLE	RESPECT SCORE
<p>FIRST NAME ASK FOR AGENCY QUESTIONS</p> <p>[name], can I see that driver's license again? It- it's showing suspended. Is that- that's you?</p> <p>↑ ↑ ↑</p> <p>DISFLUENCY NEGATIVE WORD DISFLUENCY</p>	-1.07
<p>INFORMAL TITLE ASK FOR AGENCY ADVERBIAL "JUST"</p> <p>All right, my man. Do me a favor. Just keep your hands on the steering wheel real quick.</p> <p>↑</p> <p>"HANDS ON THE WHEEL"</p>	-0.51

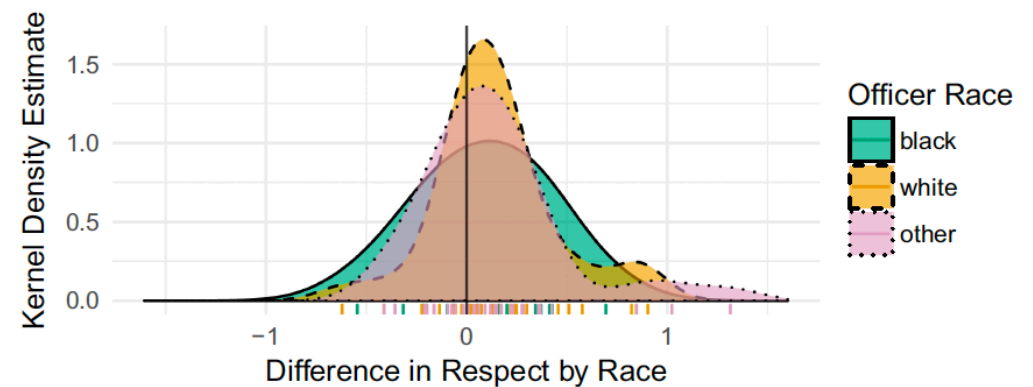
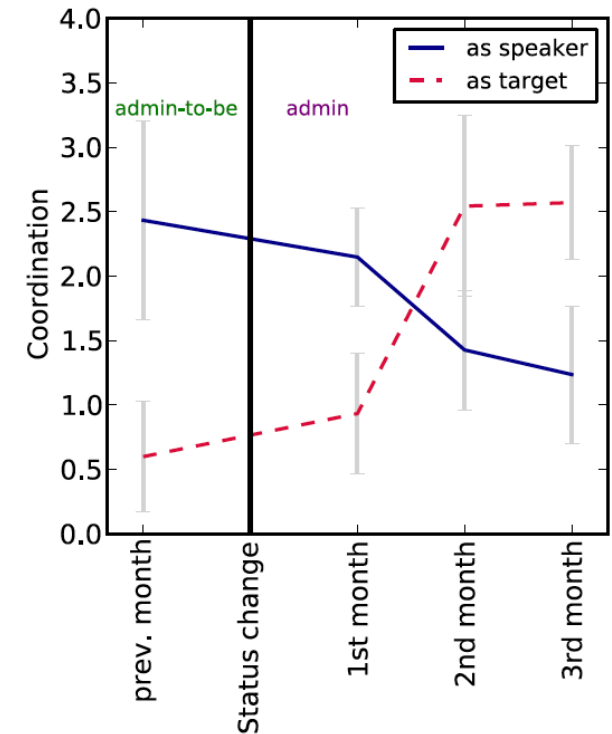


Fig. 4. Kernel density estimate of individual officer-level differences in Respect when talking to white as opposed to black community members, for the 90 officers in our dataset who have interactions with both blacks and whites. More positive numbers on the x axis represent a greater positive shift in Respect toward white community members.

Les sciences sociales concurrencées ?

Echoes of Power: Language Effects and Power Differences in Social Interaction

Cristian Danescu-Niculescu-Mizil Lillian Lee Bo Pang Jon Kleinberg
Cornell University Cornell University Yahoo! Cornell University
cristian@cs.cornell.edu, llee@cs.cornell.edu, bopang@yahoo-inc.com, kleinber@cs.cornell.edu



(b) Timed effect of **status change** (\mathcal{P})

Les sciences sociales concurrencées ?

Les imbéciles venus de la physique et de l'informatique qui ont imaginé pouvoir répondre à toutes les questions posées par les sciences sociales et les humanités avec leurs modèles de graphes exponentiels aléatoires et leurs téraoctets de corpus de textes vont finir dans le même dépotoir que les sociométristes et les physiciens sociaux des années 1930, les sociobiologistes des années 1960, les théoriciens des jeux des années 1950 et des années 1990, et ainsi de suite. (...) Ils n'ont pas vraiment d'intérêt pour les questions importantes des humanités ou des sciences sociales.

Andrew Abbott (2016)



Quel apport à la sociologie ?

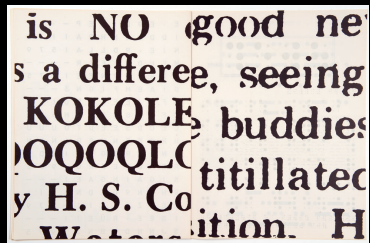
- Les plateformes contrôlent la mise en forme et l'accès aux données
- Des algorithmes opaques et étrangers aux sciences sociales
- Un ancrage social des personnes largement inconnu

→ Comment redonner de l'épaisseur sociale à ces inscriptions textuelles ?

Quel apport à la sociologie ?



ENQUÊTER À PARTIR DES TRACES TEXTUELLES DU WEB



Présentation

Jean-Philippe Cointet et Sylvain Parasio

Temporalité, régimes de participation et formes de communautés

Comprendre la dynamique d'un forum grand public autour du dépistage prénatal

Madeleine Akrich

Hostilité et prosélytisme des communautés politiques

Le militantisme politique à l'ère des réseaux sociaux

David Chavalarias, Noé Gaumont, Mazyar Panahi

La tête dans les étoiles ?

Faire sens de l'engagement dans le projet de science participative SETI@home

Elise Tancoigne et Jérôme Baudry

Comment s'élabore la mémoire collective sur le web ?

Une analyse qualitative et quantitative des pratiques d'écriture en ligne de

la mémoire de la Grande Guerre

Valérie Beaudouin

Structuration des discours au sein de Twitter durant l'élection présidentielle française de 2017

Entre agenda politique et représentations sociales

Pierre Ratinaud, Nikos Smyrniaios, Julien Figeac, Guillaume Cabanac, Ophélie Fraisier, Gilles Hubert, Yoann Pitarch, Tristan Salord, Thibaut Thonet

Comment se forment les publics d'une carte de crimes ?

Une analyse computationnelle de traces textuelles

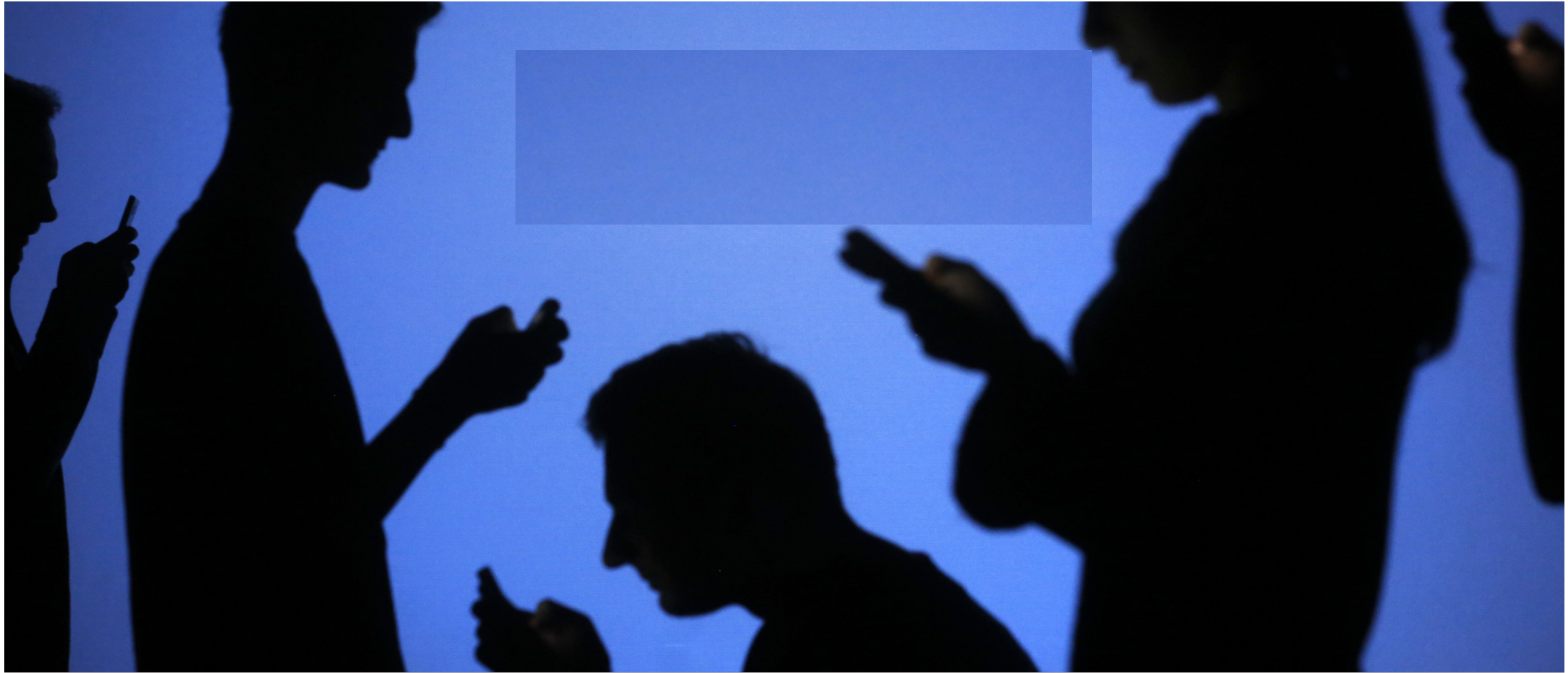
Sylvain Parasio et Jean-Philippe Cointet

Les échos du pouvoir

Effets de langage et écarts de pouvoir dans les interactions sociales

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, Jon Kleinberg

Quel apport à la sociologie ?



1. Catégoriser les locuteurs

TEMPORALITÉ, RÉGIMES DE PARTICIPATION ET FORMES DE COMMUNAUTÉS

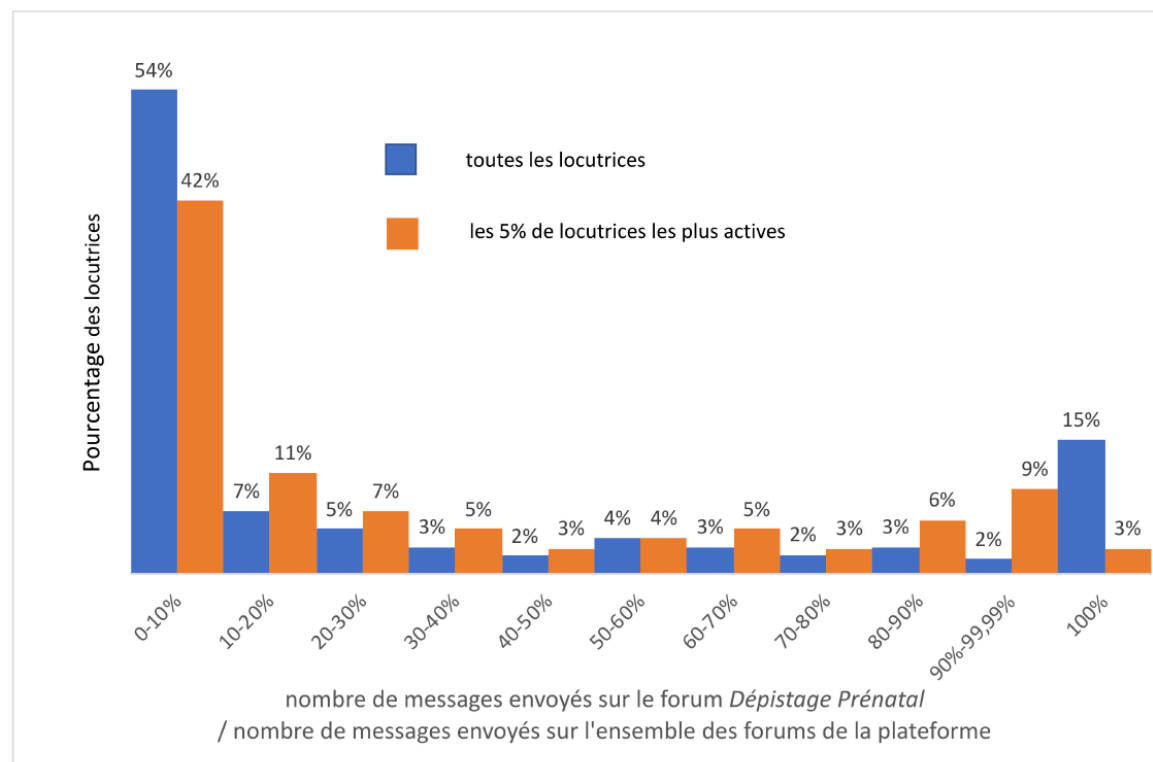
- 169 889 messages postés entre juin 2005 et juin 2018
- Analyse quali-quantitative

Comprendre la dynamique d'un forum grand public
autour du dépistage prénatal

Madeleine AKRICH¹

1. Catégoriser les locuteurs

Graphique 3. Répartition des locutrices en fonction de leur activité relative sur le forum dépistage



M. Akrich,

« Temporalité, régimes de participation et formes de communautés. Comprendre la dynamique d'un forum grand public autour du dépistage prénatal »

(2019)

- Un groupe constitué de femmes qui ne sont pas des habituées du site et sont arrivées suite à un problème lié au dépistage
- Un groupe constitué d'utilisatrices des forums du site, pour lesquelles le forum Dépistage Prénatal n'est qu'un espace possible d'échange

1. Catégoriser les locuteurs

Graphique 8. Activité des locutrices du sujet *Après l'amnio*

M. Akrich,

« Temporalité, régimes de participation et formes de communautés. Comprendre la dynamique d'un forum grand public autour du dépistage prénatal »

(2019)



1. Catégoriser les locuteurs

Do Anti-Immigrant Laws Shape Public Sentiment? A Study of Arizona's SB 1070 Using Twitter Data¹

René D. Flores
University of Washington

- corpus de 250,000 tweets relatifs à l'immigration, publiés par des résidents d'Arizona (et du Nevada) sur une période de 6 mois
- 40% des utilisateurs de Twitter mentionnent leur localisation

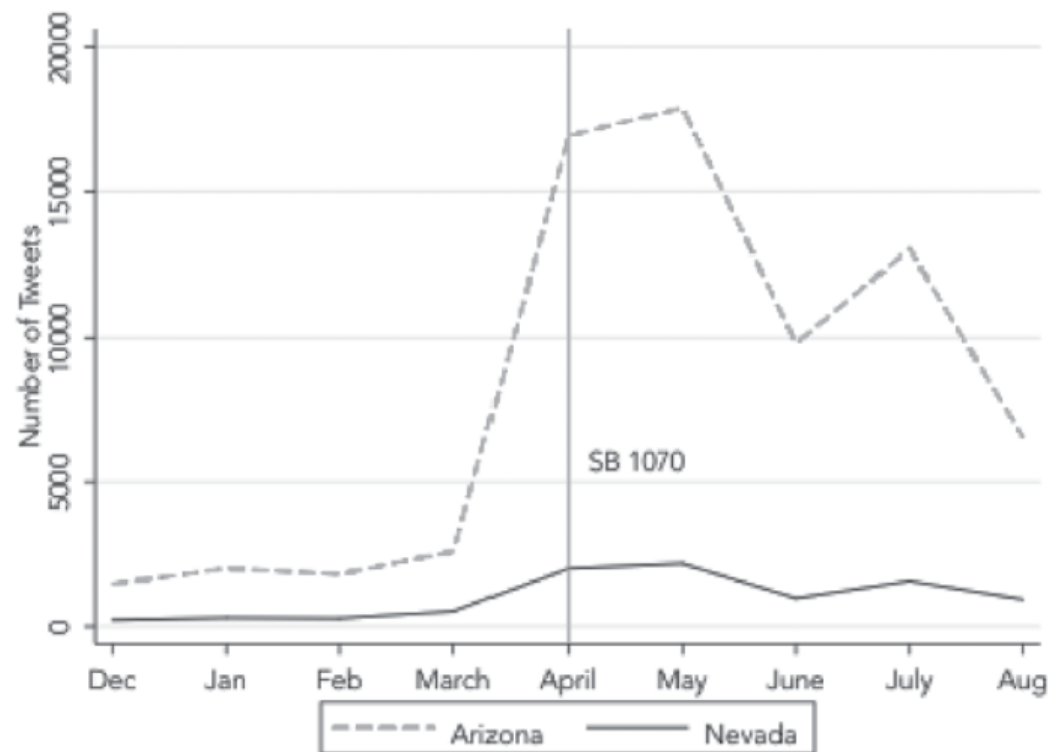


FIG. 2.—Number of Twitter messages related to immigrants per month in Arizona and Nevada (December 2010–August 2011). The vertical line on April 2010 indicates when the Arizona governor approved SB 1070.

1. Catégoriser les locuteurs

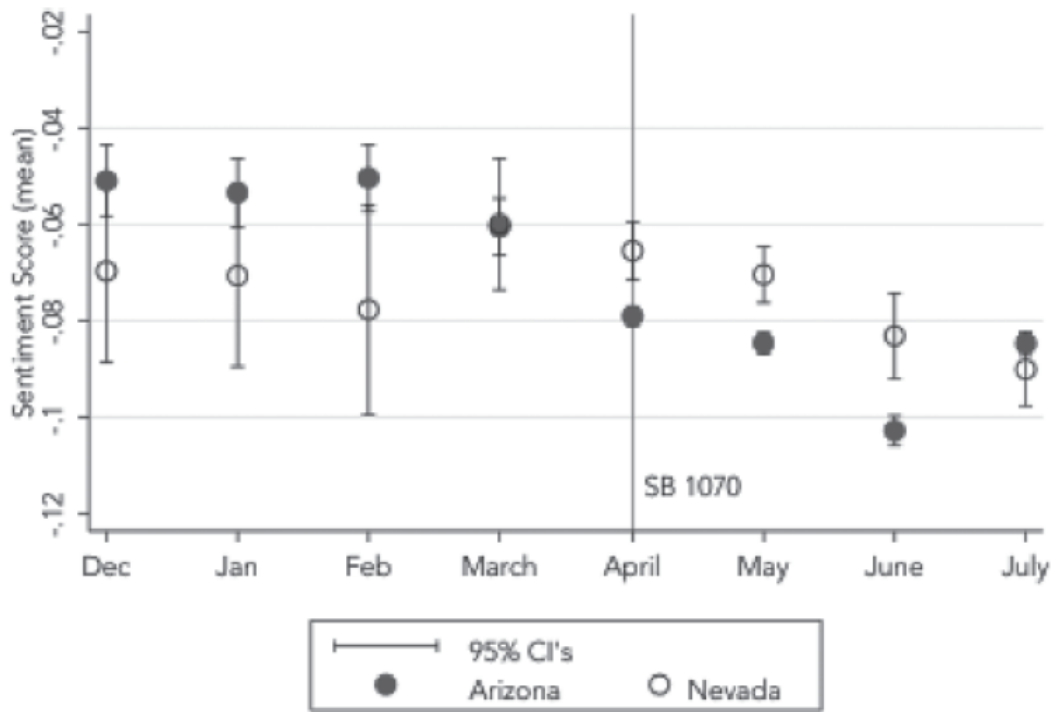


FIG. 3.—Average sentiment score of tweets about immigrants. The vertical lines represent 95% confidence intervals. The vertical line on April 2010 indicates when the Arizona governor approved SB 1070.

TABLE 6
ESTIMATED EFFECT OF SB 1070 ON NUMBER OF MESSAGES POSTED BY GROUP

	Immigrant (1)	Mexican (2)	Hispanic (3)	Asian (4)	Black (5)
Anti	36.01*** (8.88)	1.05+ (.65)	.93 (1.09)	-.05 (.36)	2.01+ (1.20)
Neutral	-11.94 (9.82)	2.66 (1.85)	.71 (2.48)	.23 (.99)	1.26 (1.17)
Pro	-.59 (3.12)	-.10 (.89)	.17 (.90)	1.39* (.54)	1.72 (1.23)

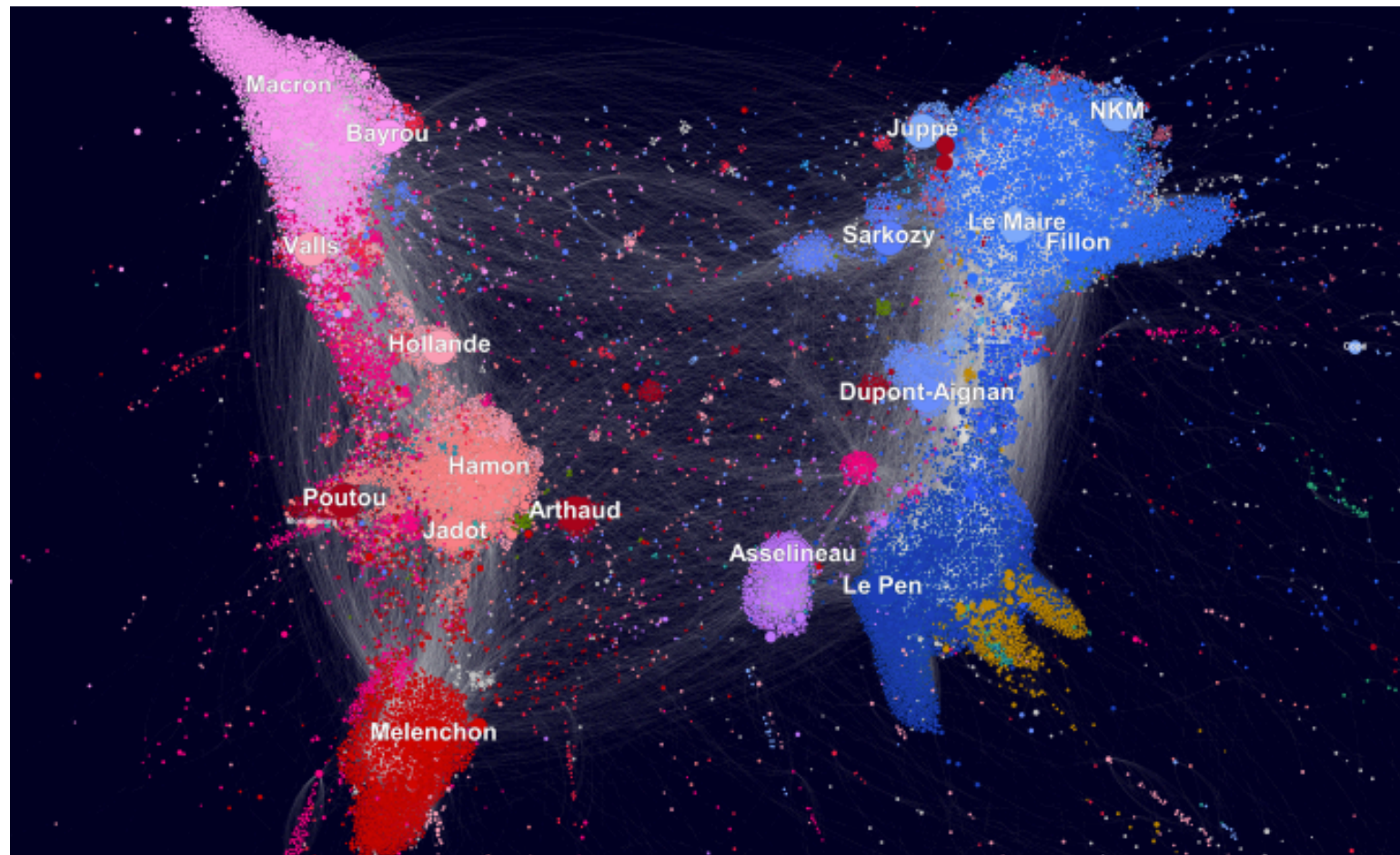
NOTE.—SEs are in parentheses. The table shows the estimated impact of SB 1070's proposal on each user group for Twitter users who posted messages before and after SB 1070 was considered. These estimates were calculated using a difference-in-difference estimation technique for each group based on coefficients in table A1. The data consist of all geocoded immigration-related tweets published in Arizona and Nevada between January and June 2010 for users who published tweets about each group before and after SB 1070 was implemented in April 2010.

- + $P < .10$.
- * $P < .05$.
- ** $P < .01$.
- *** $P < .001$.

1. Catégoriser les locuteurs

David Chavalarias,
Noé Gaumont,
Maziyar Panahi
(2019)

Communautés de la
twittersphère
politique française
pré-premier tour
2017



1. Catégoriser les locuteurs



2. Combiner les méthodes d'enquête

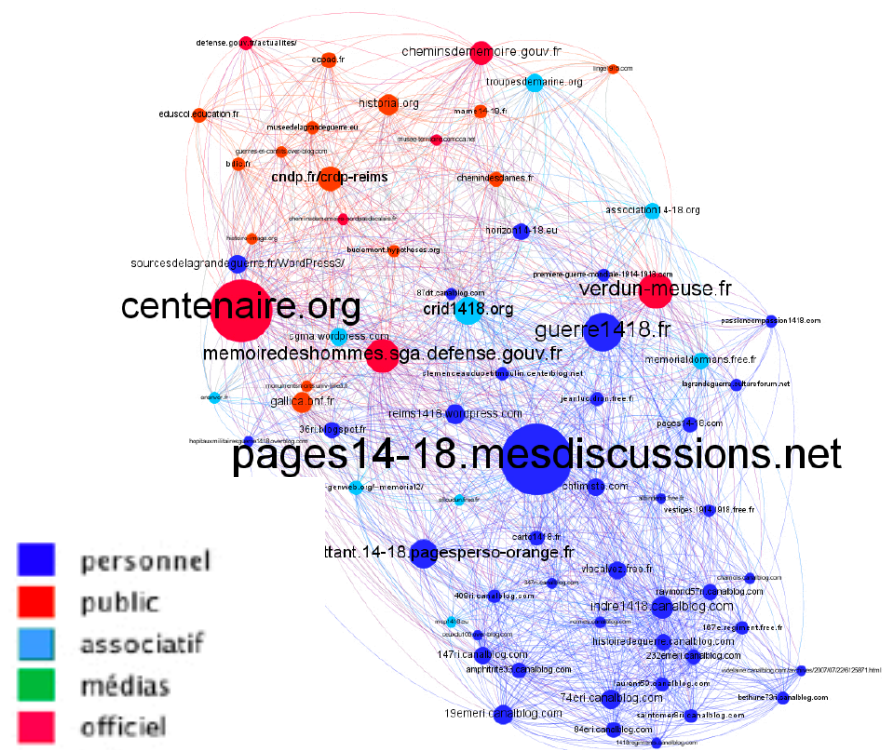
COMMENT S'ÉLABORE LA MÉMOIRE COLLECTIVE SUR LE WEB ?

Une analyse qualitative et quantitative des pratiques d'écriture en ligne de la mémoire de la Grande Guerre

Valérie BEAUDOUIN


- Corpus de 512 sites construit avec les archives du web (BNF)
- Corpus de 400 000 messages postés sur un forum (2004-2014)
- 12 entretiens avec des contributeurs

Figure 1. Carte du Web Grande Guerre (collecte de novembre 2014, sites avec degrés >30)



2. Combiner les méthodes d'enquête

Figure 3. Classification descendante hiérarchique des titres des fils de discussion sur le forum (54 639 fils entre août 2004 et janvier 2015) avec Iramuteq



classe 4 14%	classe 3 9%	classe 1 49%	classe 2 20%	classe 5 8%
nécropole, cimetière, nn, mplf, tombe, sépulture, tomber, carré, relever, photo, mam, monument, aisne, demander, militaire, marne, ossuaire, isoler, etc, canton, photographe, proposition, natif, sepulture, national, sarrebourg.	obus, douille, artillerie, casque, canon, identification, fusée, adrian, schneider, pointe, char, képi, identifier, munition, vendre, mle, chamond, balle, spécial, lebel, kepi, sabre, fusil, 37mm, ft, baïonnette, modèle.	ri, régiment, matricule, jmo, soldat, infanterie, bataillon, prisonnier, fichier, recherche, parcours, r, armée, regiment, carte, ric, citation, historique, i, bcp, uniforme, registre, officier, abréviation, territorial, médaille, bca, camp.	commémoration, centenaire, guerre, exposition, presse, grand, expo, livre, vestige, patrimoine, net, histoire, visite, journée, verdun, mémoire, bataille, musée, the, salon, mondial, conférence, cérémonie, novembre, abri, ouvrage, hommage	patrouilleur, auxiliaire, cargo, goélette, trois_mâts, navire, vapeur, armement, torpilleur, chalutier, sous_marin, dundee, dragueur, remorqueur, barque, compagnie, ii, maritime, marin, pêche, voilier, société, bateau, ex, paquebot.
Pages mémoire: nécropoles - MPLF-MDH	Collectionneurs et artillerie	Qui cherche quoi ? Soldats de la GG	Actualités 14-18, Commémoration	Marine 14-18

Entretiens :

- Attachement au travail de recherche (signalement des sources, preuves documentaires)
- Volonté de rendre visible des destins individuels
- Valorisation de l'échange et du partage
- Importance de la dimension émotionnelle de l'histoire
- Tensions avec les historiens professionnels

2. Combiner les méthodes d'enquête

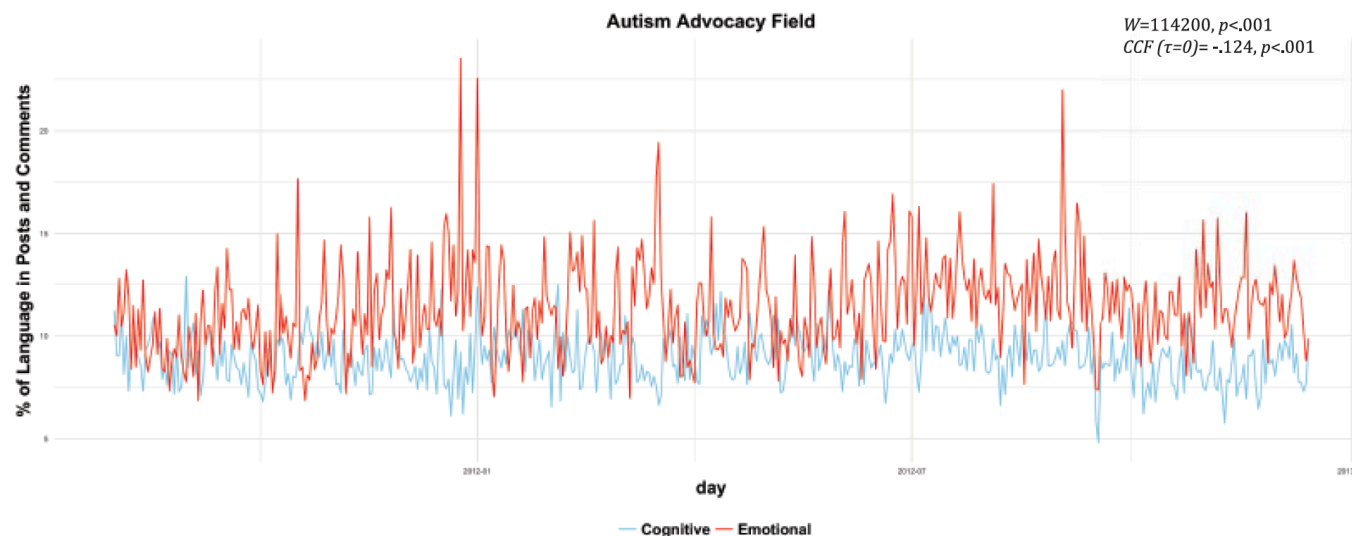
Channeling Hearts and Minds: Advocacy Organizations, Cognitive-Emotional Currents, and Public Conversation

Christopher A. Bail,^a Taylor W. Brown,
and Marcus Mann^a

American Sociological Review
2017, Vol. 82(6) 1188–1213
© American Sociological
Association 2017
DOI: 10.1177/0003122417733673
journals.sagepub.com/home/asr



- 220 associations (autisme et don d'organe)
- corpus de 223,000 posts et commentaires sur Facebook, sur une période de 18 mois



2. Combiner les méthodes d'enquête

Tactics of Advocacy Organizations

Number of Page Views Organization Received from Facebook Advertising

Organizations' Posts Discuss Influential Topic (1=yes, 0=no)

Number of Audiovisuals in Organization's Posts, Previous Day

Characteristics of Advocacy Organization's Audience

Total Number of People Who Viewed Organization's Page

Opinion Leadership (Popularity of Users who Comment on Post), Previous Day

Homophily Index (Dissimilarity of Audience Demographics)

% of Organization's Page Viewers under Age 35

% of Organization's Page Viewers Female

% of Organization's Page Viewers from Eastern U.S.

% of Organization's Page Viewers from Southwestern U.S.

% of Organization's Page Viewers from Midwestern U.S.

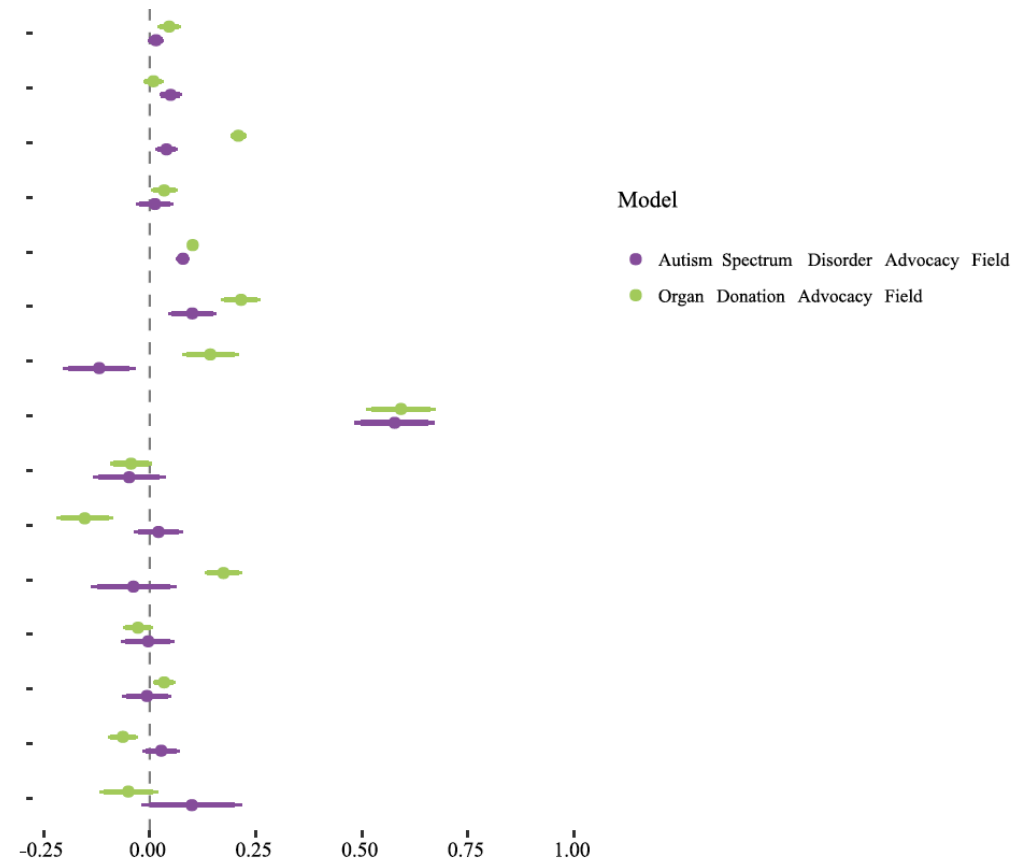
% of Organization's Page Viewers from Western U.S.

External Opportunity Structures

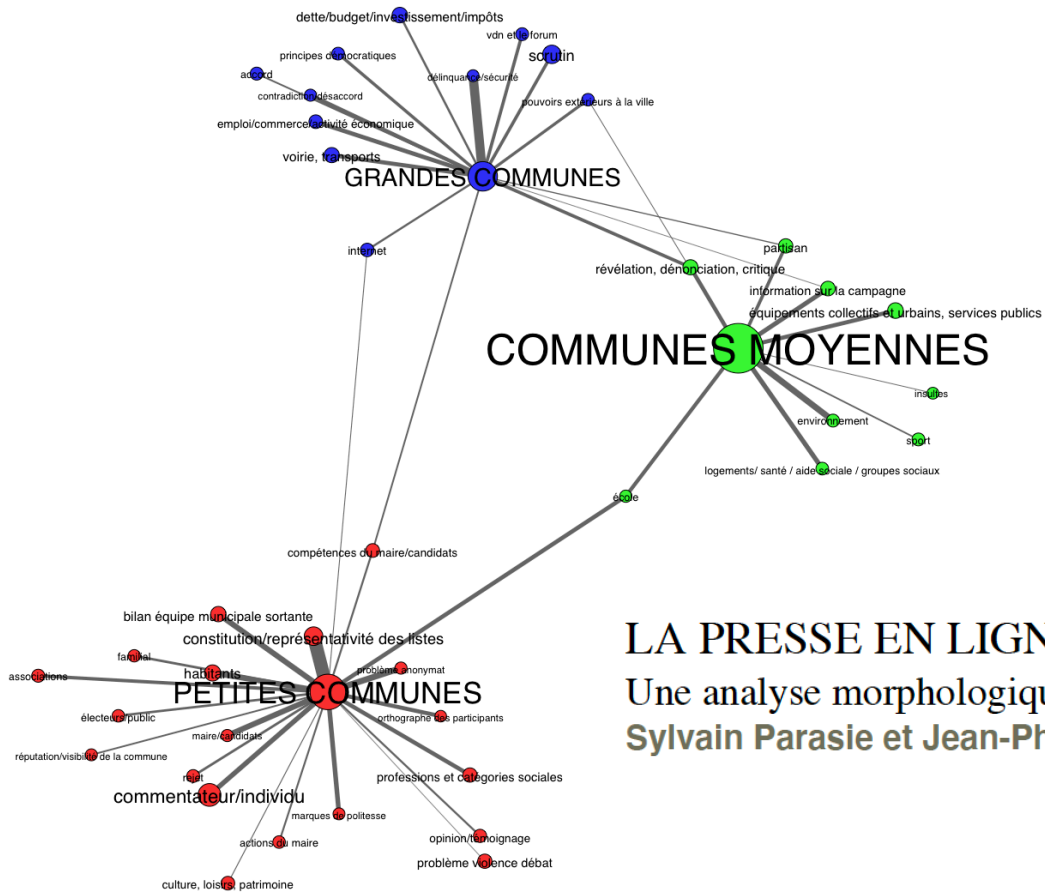
Number of News Articles about Organization, Previous Day

Number of Blog Mentions of Organization, Previous Day

Relative Volume of Google Searches, Previous Day



2. Combiner les méthodes d'enquête



LA PRESSE EN LIGNE AU SERVICE DE LA DÉMOCRATIE LOCALE

Une analyse morphologique de forums politiques

Sylvain Parasio et Jean-Philippe Cointet

2. Combiner les méthodes d'enquête

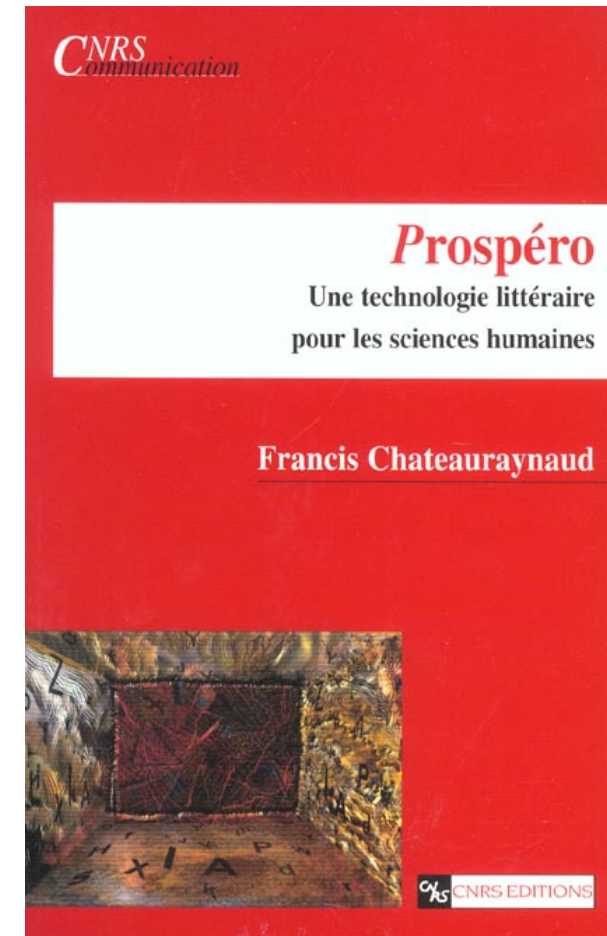
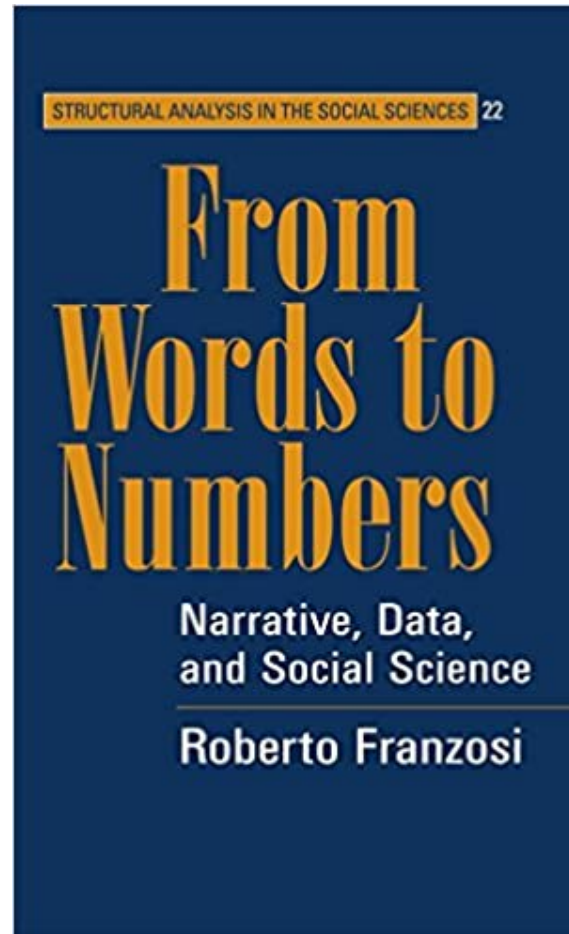


3.Saisir des actes d'énonciation

Luc Boltanski
L'Amour et
la Justice comme
compétences



folioessais



3. Saisir des actes d'énonciation

Luc Boltanski

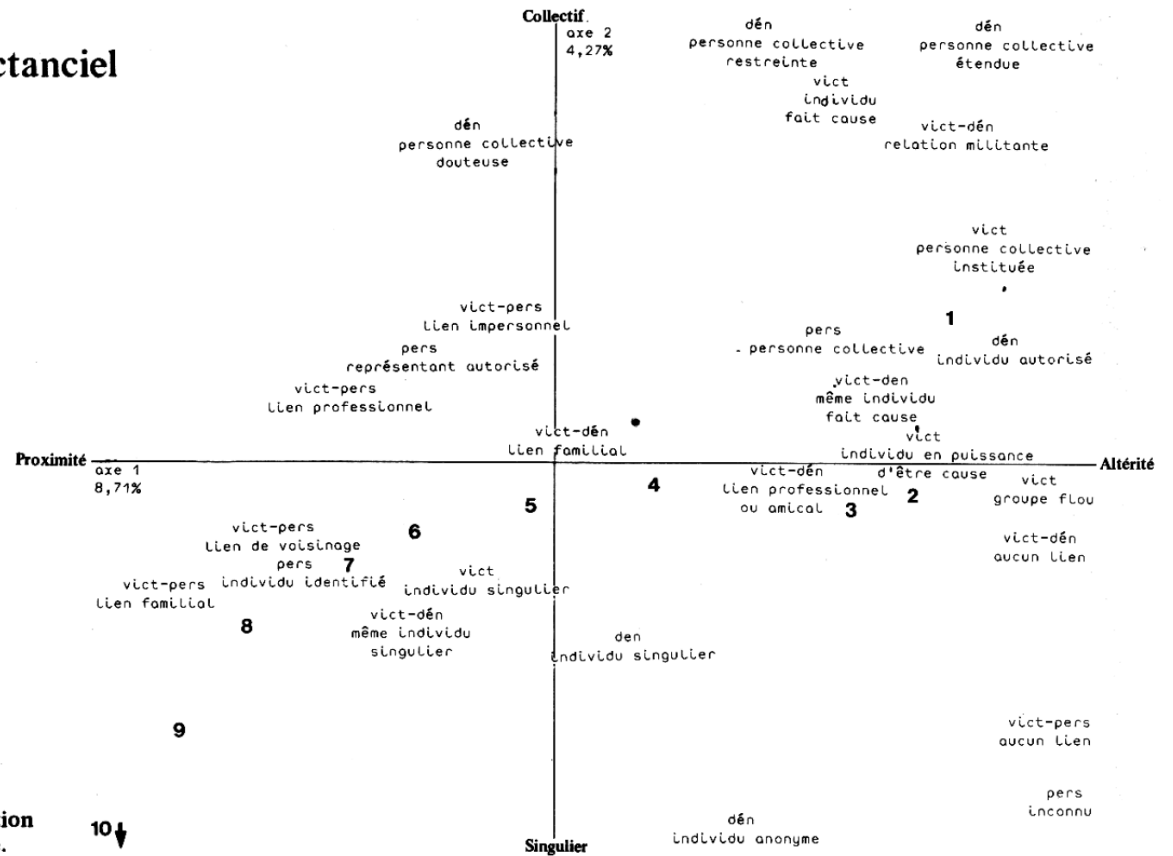
L'Amour et la Justice comme compétences



folioessais

8 Luc Boltanski

La structure du système actanciel



NB. Les chiffres représentent la projection des notes de normalité.

3. Saisir des actes d'énonciation

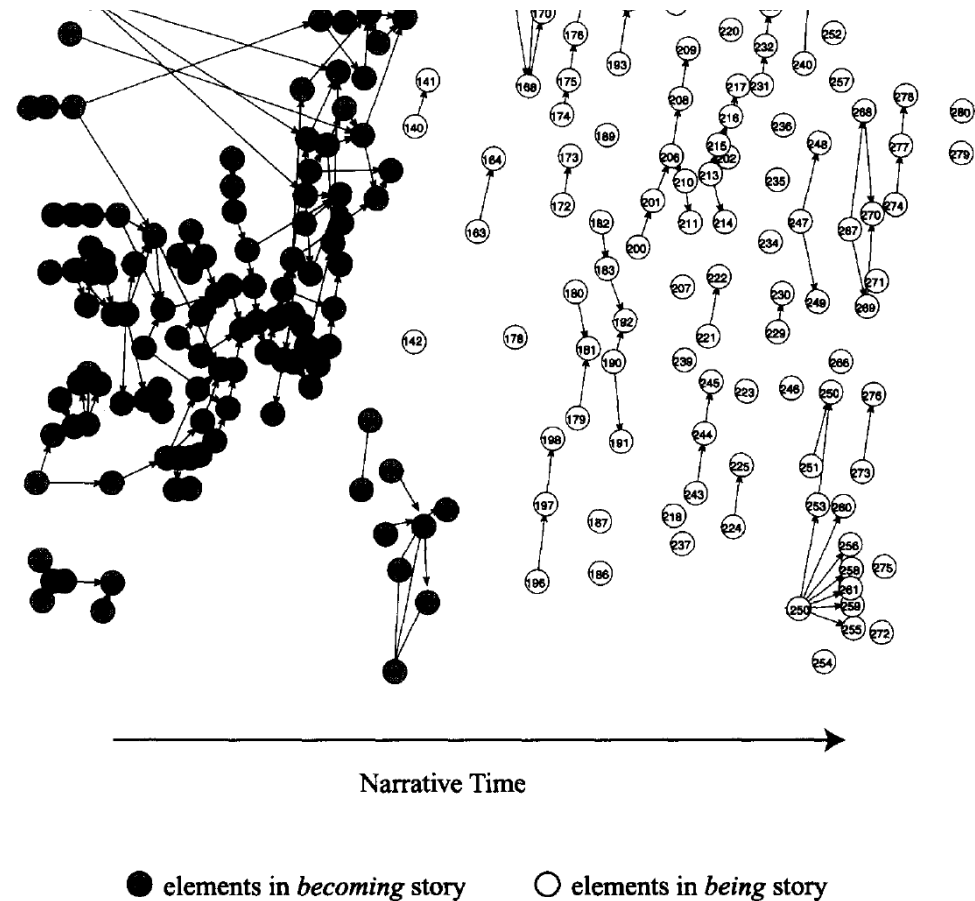
Becoming a Nazi: A model for narrative networks☆

Peter S. Bearman^{a,*}, Katherine Stovel^b

POETICS

(2000)

Structure narrative des récits de vie nazis

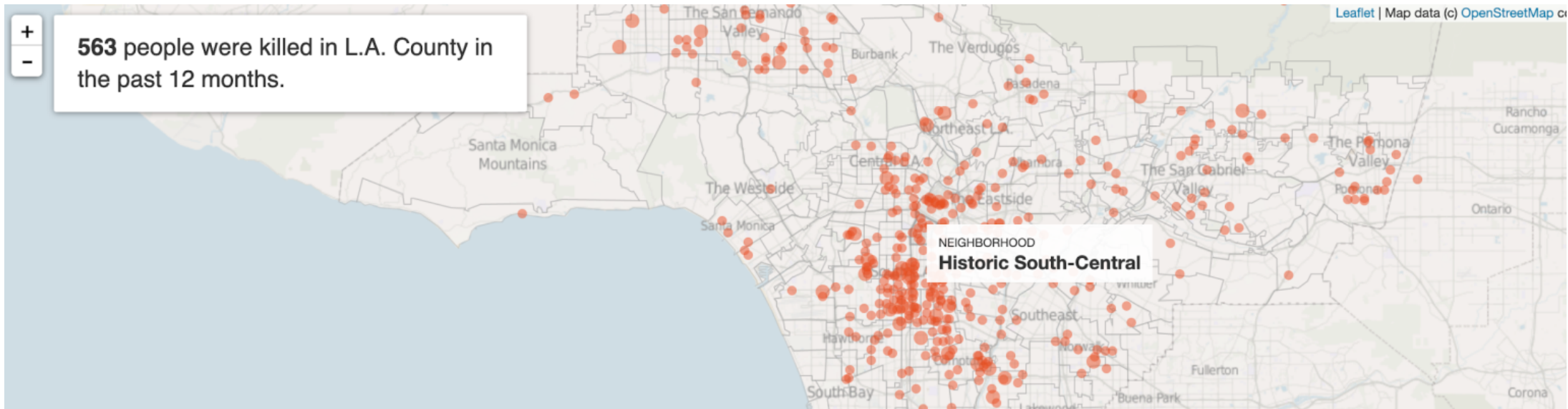


3.Saisir des actes d'énonciation

COMMENT SE FORMENT LES PUBLICS D'UNE CARTE DE CRIMES ?

Jean-Philippe COINTET
Sylvain PARASIE

Une analyse computationnelle de traces textuelles¹



3. Saisir des actes d'énonciation

“ He was a loving homie of mine and I'm going to miss him especially since he was supposed to marry my home girl Jessica I'm gonna miss him Joaquin why did you die for huh what happen that night homie plz someone explain to me what happen that night I'm confused ok

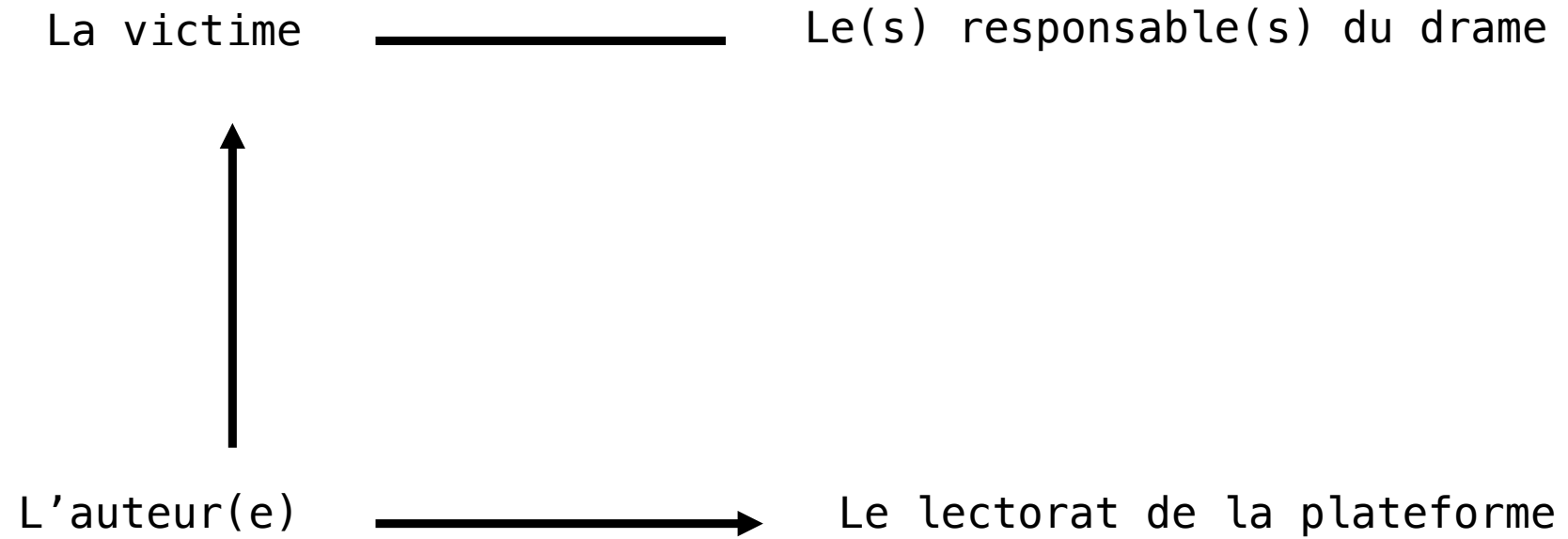
— Bella Chavez

Sept. 15, 2016 at 11:50 a.m. [↗](#)

Qui sont les auteurs?

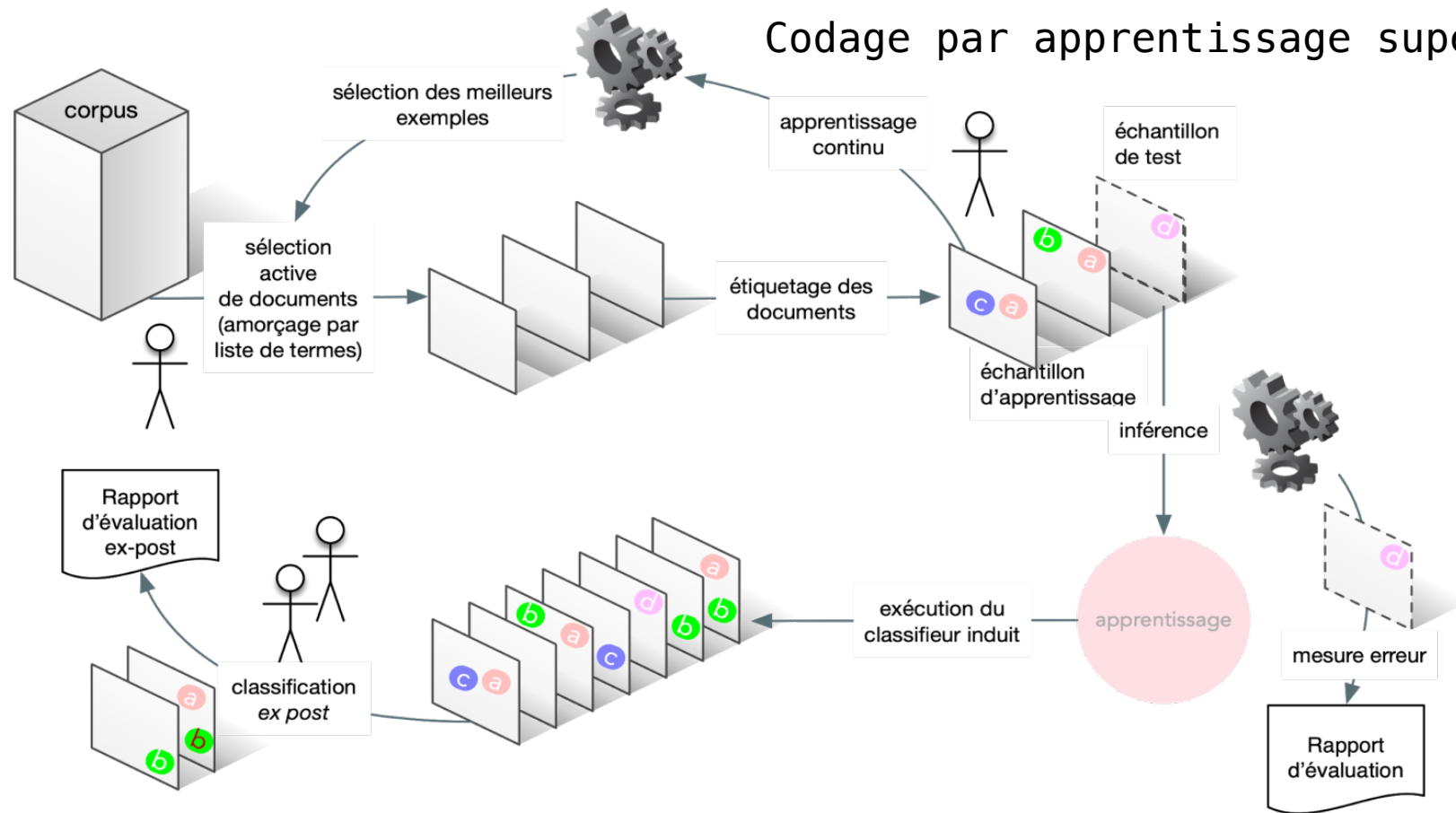
Que peut-on savoir du commentaire comme acte social ?

3.Saisir des actes d'énonciation

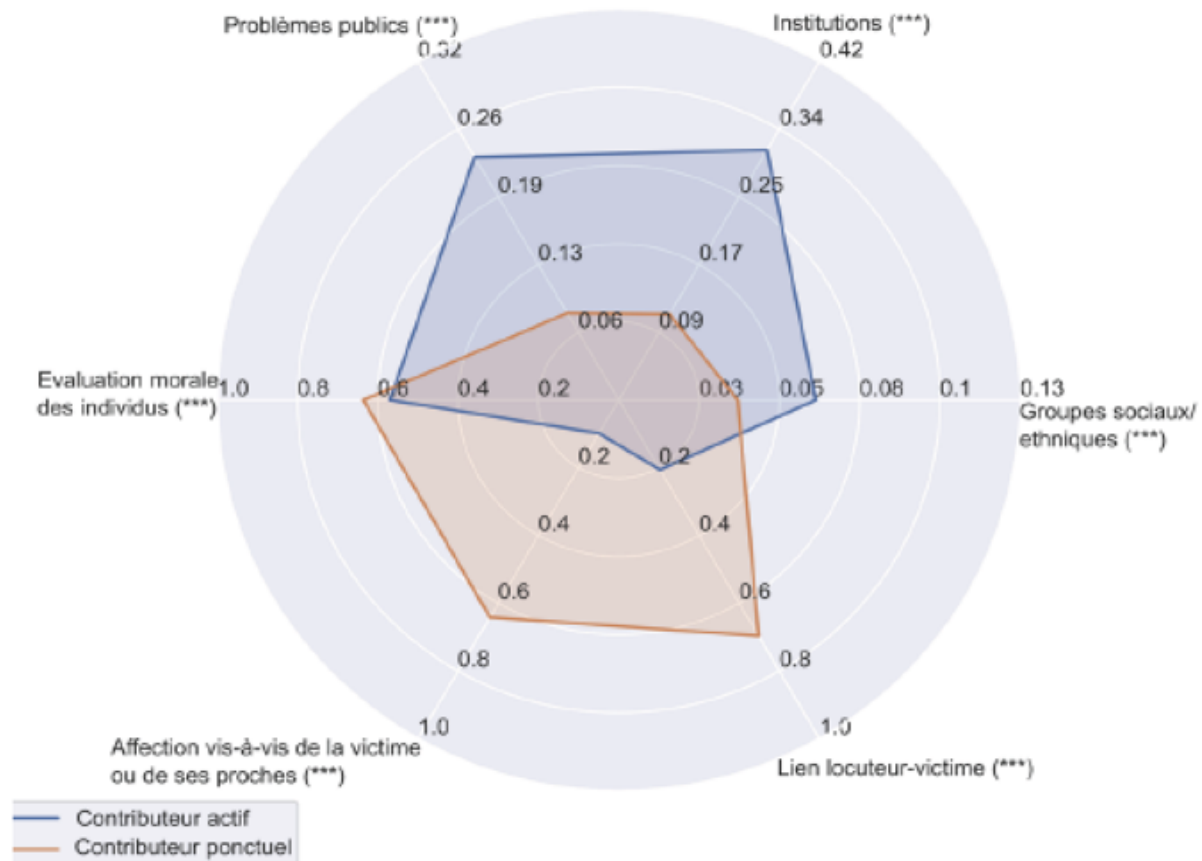


3.Saisir des actes d'énonciation

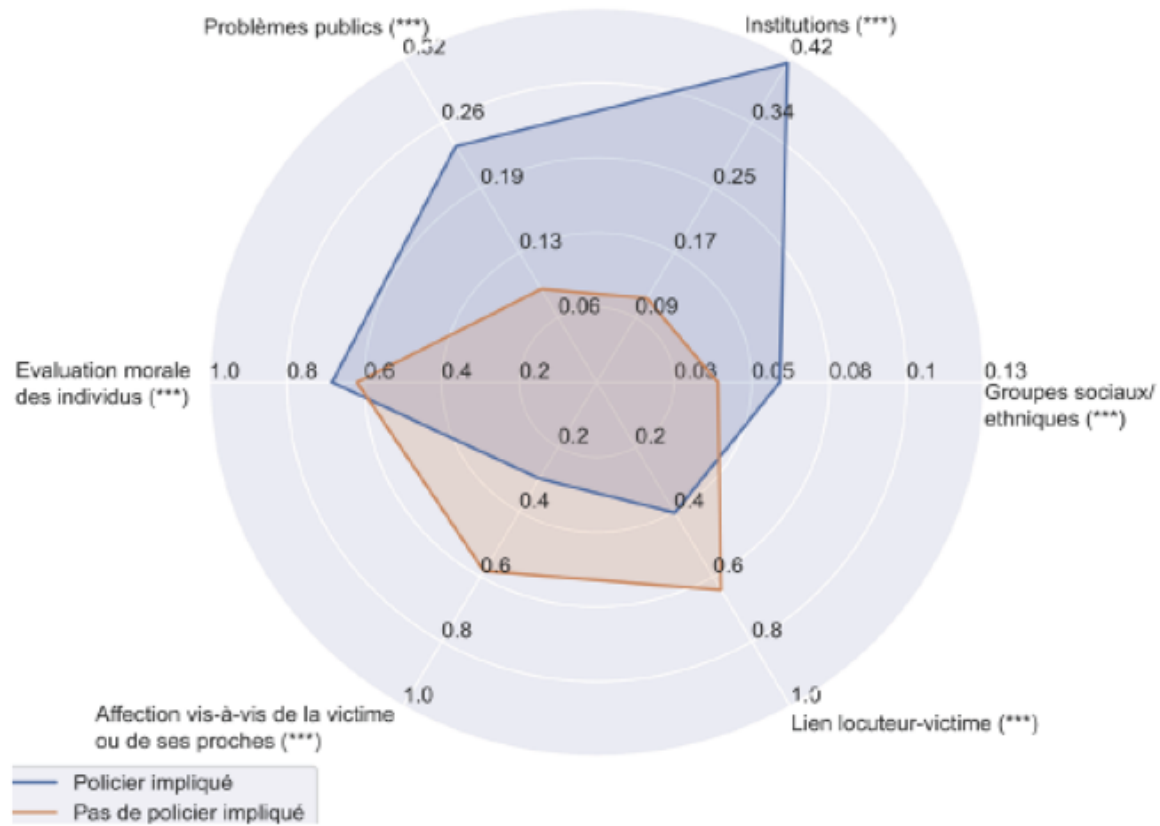
Codage par apprentissage supervisé



3. Saisir des actes d'énonciation



3.Saisir des actes d'énonciation




3.Saisir des actes d'énonciation

- Ajouter de l'extériorité aux textes
- Contrôler les « briques »
- Défendre une épistémologie qualitative

**Computational
Grounded Theory:
A Methodological
Framework**

Laura K. Nelson¹

Sociological Methods & Research
2020, Vol. 49(1) 3-42
© The Author(s) 2017
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0049124117729703
journals.sagepub.com/home/smr


Conclusion