

# Du texte aux données textuelles : quels parcours interprétatifs et quels dispositifs attentionnels ?

École internationale d'été de l'Université  
Paris-Sorbonne, 3 juillet 2015  
Clémence Jacquot – LABEX OBVIL

# Lettre-Océan

Je traverse la ville  
et je la coupe en 2

J'étais au bord du Rhin quand tu partis pour le Mexique  
Ta voix me parvient malgré l'énorme distance  
Gens de mauvaise mine sur le quai à la Vera Cruz

Les voyageurs de l'Espagne devant faire  
le voyage de Coatzacoalcos pour s'embarquer  
je l'envoie cette carte aujourd'hui au lieu

de profiter du courrier de Vera Cruz qui n'est pas sûr  
Tout est calme ici et nous sommes dans l'attente  
des événements.

Juan Aldama

Correos  
MEXICO  
4 centavos

U.S. Postage  
2 cents 2

REPUBLICA MEXICANA  
TARJETA POSTAL

Rue des Baignoilles

Vive la République  
Hoo le croquant  
Des clefs j'en ai vu mille et mille  
Vive la République  
Zut pour M. Zun  
Sur la rive gauche  
le pont d'Iéna  
Evviva il Papa  
Vive le Roi  
ta gueule mon vieux pad  
Jacques c'est  
taiz de li  
non si vous avez  
une moule tache  
un journal

BONJOUR ANOMO ANORA  
TU NE CONNAITRAS JAMAIS BIEN  
LES

# Mayas

Te souviens-tu du tremblement de terre entre 1885 et 1890  
ou coucha plus d'un mois sous la tente

BONJOUR MON FRÈRE ALBERT à Mexico

# Jeunes filles à Chapultepec

Tous saint  
Luca  
est main  
tenant  
à Poi  
tiers

et com  
ment  
j'ai brû  
le  
le  
dur  
avec  
ma  
gerce

rue Si-  
Isidore  
à la  
Havane  
cela  
n'existe  
+

LES CHAUSSURES NEUVES DU PORT

cré  
cré  
cré  
GRAMOPHONES  
z z z z z  
cré  
AUTOBUS  
r r o o o

Chirimoya

ture les voyageurs pour Chatou

les ting BIRENES to z cré  
cré mez Hoo Hou ro z cré  
cré ting Hou Haute de ou ro z  
allons circulez Mes fer Hou 300 mètres ou flag ou cré  
cré rls tion ou ou ou ting ou cré  
cré fleu sec ou ou ou  
dins de ment ge chan o ou  
jar vos de o o o o o cre cré  
cré  
cré  
cré  
cable le  
gramme  
compor  
tail 2  
mots  
EN  
SURETÉ

je me  
suis  
levé à  
2 h.  
du  
matin  
et  
j'ai  
déjà  
bu  
un  
mou  
ton

priè  
taire  
de  
5  
ou  
6  
im

Pen  
deco  
c'est  
+  
qu'un  
im  
bécile

il ap  
pelait  
l'Indien  
Hijo  
de  
la Cin  
ga  
da

A la Crème à

# La notion de « corpus »

- répertoire scientifique
- collection de données langagières  
→ échantillon du langage
- ensemble de données exploitables  
→ analyse / recherche automatique

# L'annotation de corpus

- enrichissement du texte « brut »
- ajout d'informations au corpus d'étude  
→ étiquettes, balises (XML)
- informations attribuées au texte :  
→ les *métadonnées*

# Les métadonnées

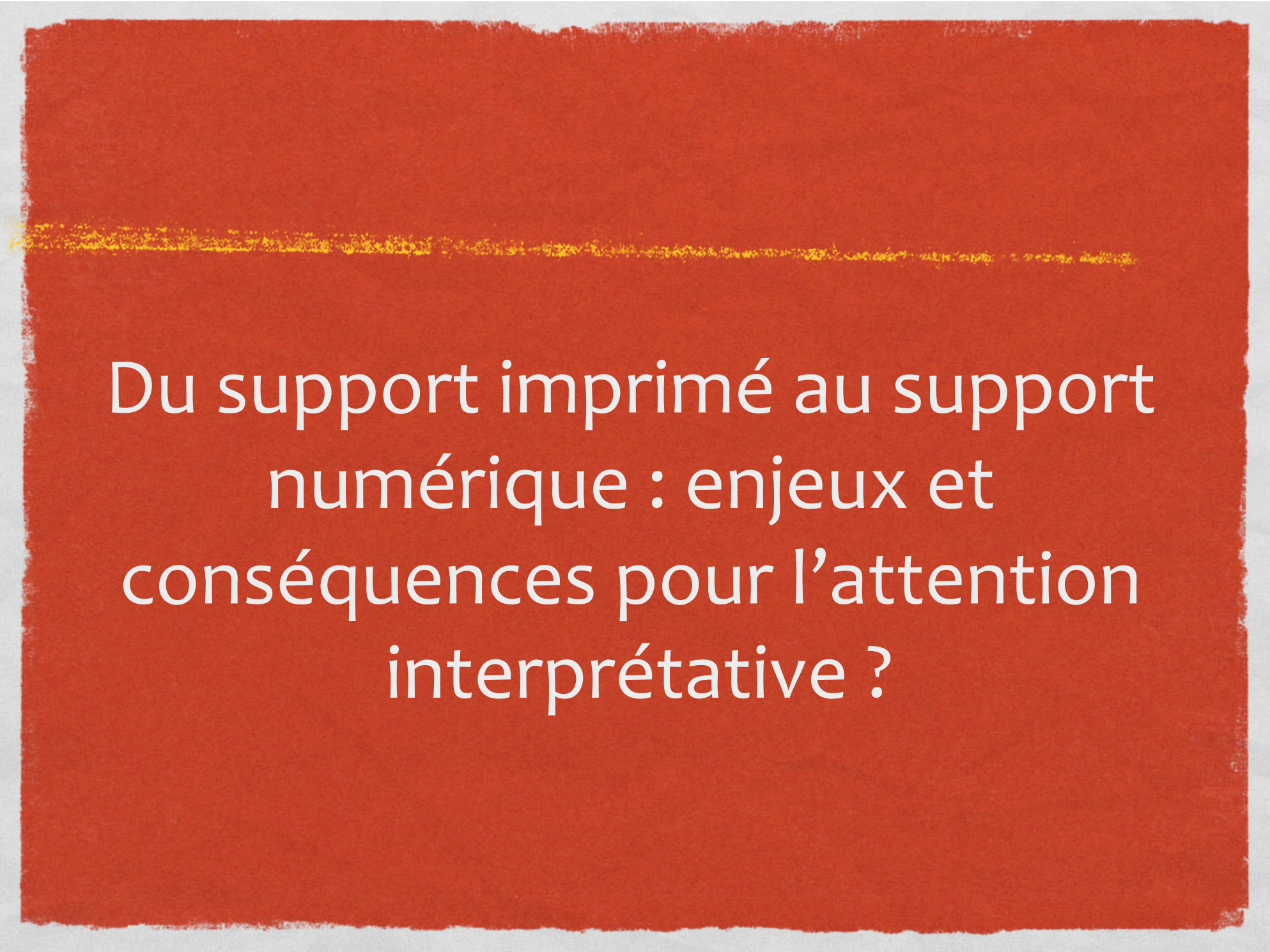
- Au niveau « macro- » :
  - Informations *structurelles* (recueil, section, titre, strophe, vers, etc.)
  - Informations *péritextuelles* (date de rédaction, de publication, support de publication, dédicace, etc.)

# Les métadonnées

- Au niveau « micro- » :
  - Informations *morpho-syntaxiques* (catégorie grammaticale d'un mot ou d'une locution)
  - Lemmatisation (ex. « clef » ← *clé, clés, clefs, clef*)
  - Informations *sémantiques* (champs lexicaux, champs sémantiques, entités nommées, etc.)

# Lecture et interprétation

- Deux rapports distincts au texte
- L'attention interprétative : une disponibilité face à l'objet
- L'interprétation stylistique : le « filtrage » du style
- Attention et appréhension du support : des conditions de l'interprétation

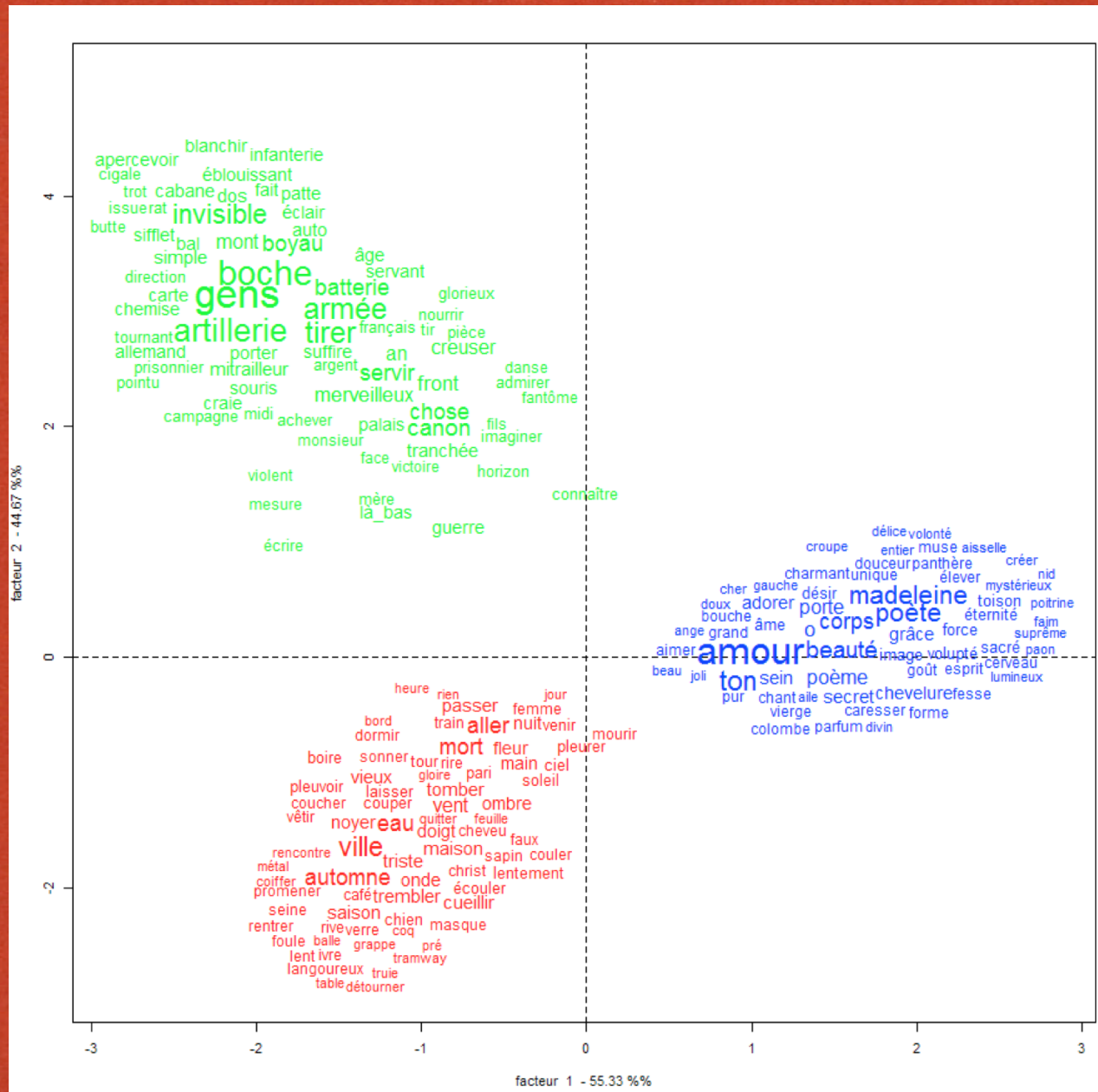
The background is a vibrant red with a textured, brush-stroke-like appearance. A horizontal yellow line, also with a textured, brush-stroke-like appearance, runs across the upper portion of the red area. The text is centered and written in a clean, white, sans-serif font.

Du support imprimé au support  
numérique : enjeux et  
conséquences pour l'attention  
interprétative ?









Visualisation en AFC des mots lexicaux des œuvres poétiques d'Apollinaire

# Support numérique et attention portée au texte

- Construction de corpus et projections interprétatives
  - Les transformations du texte initial
- Les outils de traitement numérique : un *knowledge design* ?

# Bibliographie

- Auroux, S. (1993), *La Révolution technologique de la grammatisation*, Paris, Mardaga.
- Bachimont, B. (2004), *Arts et Sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Mémoire d'HDR, U.T.C., [URL : [http://www.utc.fr/~bachimon/Livresettheses\\_attachments/HabilitationBB.pdf](http://www.utc.fr/~bachimon/Livresettheses_attachments/HabilitationBB.pdf)].
- Bachimont, B. (2014) « Le nominalisme et la culture : questions posées par les enjeux du numérique » in Stiegler, B. [dir.], *Digital Studies. Organologie des savoirs et technologies de la connaissance*, Paris, FYP Éditions, p. 63-78.
- Citton, Y. (2010), *L'Avenir des Humanités. Économie de la connaissance ou cultures de l'interprétation ?*, Paris, La Découverte.
- Cormerais, F. (2014), « Humanités digitales et (ré)organisation du savoir », in : Le Deuff, O. [dir.], *Le Temps des humanités digitales. La mutation des sciences humaines et sociales*, Paris, FYP Éditions, p. 129-142.
- Ghitalla, F., Boullier, D., Gkouskou-Giannakou, P., Le Douarin, L. et Neau, A. (2003), *L'Outre-lecture. Manipuler, [s']appropriier, interpréter le Web*, Paris, Bibliothèque Centre Pompidou.
- Guiraud, P. (1960), *Problèmes et méthodes de la statistique linguistique*, Paris, PUF.
- Hayles, K. (2007), « Hyper and Deep Attention : the Generational Divide in Cognitive modes », *Profession*, p. 187-199 [URL : [http://www.english.ufl.edu/da/hayles/hayles\\_hyper-deep.pdf](http://www.english.ufl.edu/da/hayles/hayles_hyper-deep.pdf)].
- Manovich, L. (2001), *The Language of New Media*. Cambridge, Massachusetts, London, England, The MIT Press.
- Mayaffre, D. (2002), « Les corpus réflexifs : entre architextualité et hypertextualité », in *Corpus*, n° 1, [URL : <http://corpus.revues.org/index11.html>].
- Moretti, F. (2005), *Graphs, Maps, Trees : Abstract Models for Literary History*, Londres, Verso.
- Muller C. (1993), *Principes et méthodes de statistique lexicale*, Paris, Champion.
- Rastier, F. (2001), *Arts et sciences du texte*, Paris, PUF.
- Rastier, F. (2004), « Enjeux épistémologiques de la linguistique de corpus », *Texte !* [URL de référence : [http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Enjeux.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html)].
- Rastier, F. (2011) *La Mesure et le grain. Sémantique de corpus*, Paris, Champion.
- Schaeffer, J.-M. (2011), « Styles attentionnels et relation esthétique », in Jenny, L. [dir.], *Le Style en acte. Vers une pragmatique du style*, Genève, Mētis Presses, p. 138-149.
- Stiegler, B. (2004), *De la misère symbolique, I. L'Époque hyperindustrielle*, Paris, Gallilée.
- Stiegler, B. (2012), *Prendre soin de la jeunesse et des générations*, Paris, Flammarion.
- Stiegler, B. (2014) « Pharmacologie de l'épistèmè numérique », in Stiegler, B. [dir.], *Digital Studies. Organologie des savoirs et technologies de la connaissance*, Paris, FYP Éditions, p. 13-26.
- Stiegler, B. (2014), « L'attention, entre économie restreinte et individuation collective », in : Citton, Y. [dir.] (2014), *L'Économie de l'attention. Nouvel horizon du capitalisme ?*, Paris, La Découverte, p. 123-135.